Trajectory Seriation via Spectral Tangent Alignment and Global Embedding

Zhixin Zhou [1] **Navin Souda** [2] **Arash A. Amini** [2]

Abstract

We study linear seriation from noisy point clouds: given samples from an unknown one-dimensional curve embedded in $\mathbb{R}^d$, recover their latent order along the curve, up to reversal. Unlike much of the seriation literature, which starts from a precomputed similarity matrix, this setting retains ambient coordinates and therefore local geometric information. We propose STAGE, a geometric ordering method that estimates local tangent directions by neighborhood PCA, resolves their sign ambiguity through a graph-wide $\mathbb{Z}_2$ synchronization step, and constructs signed local displacement estimates by projecting ambient differences onto the oriented tangents. These increments are then integrated into a global scalar embedding through an inhomogeneous least-squares problem, equivalently a Laplacian linear system with a nonzero right-hand side, and the final order is obtained by sorting the embedding. We prove a finite-sample Kendall's τ recovery bound that makes explicit the roles of curvature, noise, neighborhood scale, sampling density, and graph connectivity. Empirically, STAGE gives accurate and fast order recovery on high-dimensional synthetic curves, compares favorably with spectral seriation, 1D UMAP, 1D t-SNE, Recanati's method, LTSA, and SABRE, and produces meaningful pseudotime orderings on single-cell RNA-seq datasets.

1. Problem Formulation

We address the problem of linear seriation: recovering the intrinsic order of points sampled from a one-dimensional manifold (a curve) embedded in a higher-dimensional Euclidean space. Let $\mathcal{M} \subset \mathbb{R}^d$ be an unknown smooth one-dimensional manifold. We assume $\mathcal{M}$ can be parameterized by an injective function $\gamma : [0,1] \to \mathbb{R}^d$, where the parameter $\theta \in [0,1]$ represents the intrinsic order along the curve. We further assume the curve has non-zero speed, i.e., $\|\gamma'(\theta)\|_2 > 0$ for all $\theta \in [0,1]$. The points $\gamma(0)$ and $\gamma(1)$ define the manifold's endpoints.

We are given a set of n observations, denoted $\mathscr{X}_n = \{x_1, x_2, \ldots, x_n\} \subset \mathbb{R}^d$. Each observation x_i is generated from an unknown underlying parameter $\theta_i \in [0,1]$ and corrupted by additive noise:

$$x_i = \gamma(\theta_i) + \epsilon_i, \quad i = 1, \ldots, n. \tag{1}$$

Here, the parameters $\{\theta_i\}_{i=1}^n$ are distinct and their values are unknown. The noise terms $\{\epsilon_i\}_{i=1}^n$ are assumed to be i.i.d. zero-mean sub-Gaussian random vectors in $\mathbb{R}^d$.

The objective of linear seriation is to estimate the unknown ordering of these latent parameters $\{\theta_i\}$. Since the inherent directionality of the manifold (i.e., which endpoint corresponds to $\theta = 0$ versus $\theta = 1$) is typically not identifiable from the data $\mathscr{X}_n$ alone, we aim to find a permutation ϕ of the indices $\{1, \ldots, n\}$ such that the parameters indexed by ϕ are consistently ordered, i.e., either

$$\begin{aligned} &\theta_{\phi(1)} < \theta_{\phi(2)} < \cdots < \theta_{\phi(n)}, \quad \text{or} \\ &\theta_{\phi(1)} > \theta_{\phi(2)} > \cdots > \theta_{\phi(n)}. \end{aligned} \tag{2}$$

Alternatively, the goal can be viewed as assigning a real-valued score $s(x_i)$ to each observation x_i such that the order of these scores reflects the true order of the underlying parameters $\{\theta_i\}$, or its reverse. That is, $s(x_i)$ should be monotonically increasing or monotonically decreasing with respect to the true parameter values θ_i.

2. Overview of Related Work and Our Method

The problem of seriation, which seeks to recover a latent ordering among a set of items based on pairwise measurements, has a rich history and diverse applications. Originating in archaeology for the chronological ordering of artifacts (Robinson, 1951), its principles are now applied in fields such as sparse matrix computations for envelope reduction (Barnard et al., 1993), bioinformatics for tasks like *de novo* genome assembly (Garriga et al., 2008; Meidanis et al., 1998; Recanati et al., 2017), and network analysis for time synchronization (Elson et al., 2004).

[1]Alpha Benito Research, Irvine, USA [2]Department of Statistics, University of California, Los Angeles, USA. Correspondence to: Navin Souda <nvsouda@ucla.edu>.

Proceedings of the 43rd International Conference on Machine Learning, Seoul, South Korea. PMLR 306, 2026. Copyright 2026 by the author(s).

Many existing seriation methods operate on a pre-defined or computed $n \times n$ matrix A, representing pairwise similarities (or dissimilarities) between n items. The underlying assumption is often that this matrix, when items are correctly ordered, exhibits a specific structure, most notably the Robinson property. A matrix has Robinson property if whenever we move away from the diagonal along any row or column, the values decrease (cf. Definition 1.1 and 1.2 in (Recanati et al., 2018) and (Fogel et al., 2013)). The challenge then lies in finding the permutation ϕ that transforms the observed, permuted similarity matrix into (or close to) this ideal Robinson form. However, when the original data consists of points in a metric space (e.g., $\mathbb{R}^d$), deriving such a global similarity matrix can lead to a loss of finer geometric information inherent in the point cloud coordinates (cf. Appendix A).

2.1. Spectral and Affinity-Based Methods for Seriation

A widely adopted class of methods for seriation from such similarity matrices are spectral techniques. These methods typically construct a graph Laplacian from the similarity matrix A. For linear seriation, the seminal work by Atkins et al. (Atkins et al., 1998) demonstrated that the Fiedler vector—the eigenvector corresponding to the second smallest eigenvalue of the graph Laplacian $L_A = \text{diag}(A\mathbf{1}) - A$—provides an ordering of the items. Sorting the components of this Fiedler vector yields the estimated seriation, known to exactly recover the order for ideal Robinson matrices. Extensions of spectral methods also address circular seriation. For instance, Coifman et al. (Coifman et al., 2008) utilized the first two non-trivial eigenvectors of a normalized graph Laplacian to embed data points onto a circle, with the ordering then determined by their angles. The theoretical underpinnings of these methods often connect the discrete graph Laplacian to continuous Laplace-Beltrami operators on underlying manifolds (Belkin & Niyogi, 2003; Coifman & Lafon, 2006; Singer, 2006). Other dimensionality reduction techniques like Multidimensional Scaling (MDS) (Kruskal & Wish, 1978) and kernel PCA (Schölkopf et al., 1997) have also been explored for embedding data in ways that might preserve latent ordering. An overview of spectral clustering, which shares methodology, can be found in (Von Luxburg, 2007).

2.2. Statistical Approaches and Robustness in Noisy Seriation

Real-world data often presents noisy similarity measurements, where the observed matrix A is a perturbation of an ideal structured matrix F_ϕ. While spectral methods have proven effective, their robustness to noise, especially when F_ϕ deviates from simple structures, has been a subject of investigation (Giraud et al., 2023; Natik & Smith, 2021). Recent research has focused on developing algorithms with provable guarantees in noisy settings, analyzing specific matrix classes (Cai & Ma, 2023), or studying seriation within latent space models (Janssen & Smith, 2022). These analyses often aim to characterize error rates (Flammarion et al., 2019) and establish conditions for reliable recovery of the latent permutation ϕ.

2.3. Our Approach: Geometric Seriation Directly from Point Cloud Data

Motivated by the potential information loss in methods relying solely on pre-computed global similarity matrices, our work, STAGE (Spectral Tangent Alignment and Geometric Embedding), directly leverages the geometric structure of the original observed data points $\{x_i\}_{i=1}^n \subset \mathbb{R}^d$. Instead of first collapsing the point cloud information into a global similarity matrix, our method proceeds by:

1. **Estimating Local Manifold Geometry:** We employ local Principal Component Analysis (PCA) within neighborhoods of each data point x_i to estimate the local tangent direction $\widehat{v}_i$ of the underlying 1D manifold. This approach is inspired by techniques in manifold learning that infer local structure from point coordinates (Roweis & Saul, 2000).
2. **Global Orientation of Tangents:** A spectral method is then used to consistently orient these local tangent vectors across the manifold, yielding $\widehat{s}_i\widehat{v}_i$.
3. **Constructing a Geometrically-Informed Global Embedding:** The core of our distinction lies in using these oriented local geometric features ($\widehat{s}_k\widehat{v}_k$) and the original point coordinates (x_k) to directly formulate the objective for a global 1D embedding y. This is achieved by solving a least-squares problem, such as minimizing $\sum_{i,j:\|x_i-x_j\|_2\le r}(y_j - y_i - \widehat{W}_{ij})^2$, where $\widehat{W}_{ij} = \frac{1}{2}(\widehat{s}_i\widehat{v}_i + \widehat{s}_j\widehat{v}_j)^\intercal(x_j - x_i)$.

This direct integration of local, coordinate-based geometric estimates into the global ordering objective allows our method to be more adaptive to the specific geometry of the data manifold. While many manifold learning techniques (e.g., Laplacian Eigenmaps (Belkin & Niyogi, 2003), Diffusion Maps (Coifman & Lafon, 2006)) also start from point coordinates, they typically first construct a global similarity graph (e.g., using a heat kernel $W_{kl} = \exp(-\|x_k - x_l\|^2/t)$) and then apply spectral methods to this graph. Our approach defers the aggregation of local information until the formulation of the final embedding objective, potentially preserving finer geometric details relevant for accurate seriation. The work by Recanati et al. (Recanati et al., 2018) also explores enhancing latent orderings using multidimensional Laplacian embeddings and refining similarities based on local fits within the resulting embedding. Issartel et al. (Issartel et al.,

2024) provide a minimax optimal algorithm operating on noisy similarity matrices by aggregating local bisections. Among manifold learning methods, LTSA (Zhang & Zha, 2004) and vector-valued diffusion methods such as VDM (Singer & Wu, 2012) and LPD (Wolf & Averbuch, 2013) are particularly relevant because they also begin with local PCA and then aggregate local tangent information globally. However, these methods are designed for general manifold embedding or distance construction rather than for recovering a globally consistent 1D ordering. In particular, they do not explicitly resolve the orientation ambiguity of 1D tangent estimates and they do not construct the signed projected increments $\widehat{W}_{ij}$ that form the regression targets in STAGE. This distinction is central: STAGE is specialized to the case where the latent structure is believed to be one-dimensional and the goal is seriation rather than generic visualization or manifold parametrization. Appendix E.3 gives a more detailed comparison with LTSA, VDM, and LPD/PTE, including the distinction between STAGE's sign-synchronization matrix $\widehat{Q}$ and the projection blocks used in LPD. The robustness of our method stems from the local averaging inherent in PCA and the global consistency enforced by the spectral orientation and the final embedding stage.

3. The STAGE Algorithm

Let $G_r(\mathscr{X}_n)$ denote the standard r-neighborhood (a.k.a. ε-) graph of the sample $\mathscr{X}_n$, where

$$i \sim j \iff \|x_i - x_j\| \le r.$$

We let $A = (A_{ij}) = 1\{i \sim j\}$ be the adjacency matrix of $G_r(\mathscr{X}_n)$ and $L = D - A$, its (combinatorial) Laplacian. We let $\lambda_2(L)$ denote the smallest nonzero eigenvalue of L, a.k.a. algebraic connectivity of the graph, and let $d_i = \sum_j A_{ij}$ denote the degree of node i.

Our method proceeds in four stages: first, estimate local tangent vector at each point, using its neighbors in $G_r(\mathscr{X}_n)$; second orient them globally using a spectral method; third, compute a 1D embedding that aligns with these oriented local structures via least squares; fourth, derive the final order from the embedding. Below, we detail each step.

3.1. Local Tangent Estimation

The first stage of our algorithm estimates the local tangent direction $\widehat{v}_i \in \mathbb{R}^d$ at each point $x_i \in \mathscr{X}_n$. This is achieved by performing Principal Component Analysis (PCA) on the set of its neighbors within the r-neighborhood graph $G_r(\mathscr{X}_n)$.

For each point $x_i \in \mathscr{X}_n$, let d_i be its degree in $G_r(\mathscr{X}_n)$ (i.e., the number of points x_j such that $\|x_j - x_i\|_2 \le r$). If $d_i < d_{\min}$ (where $d_{\min} \ge 2$ is a minimum required number of neighbors, e.g., $d_{\min} = 2$ or slightly higher for stability), the local tangent cannot be reliably estimated for x_i. For all other points, we compute:

1. The local sample mean:

$$\widehat{\mu}_i = \frac{1}{d_i} \sum_{j: j \sim i} x_j. \tag{3}$$

2. The local sample covariance matrix:

$$\widehat{\Sigma}_i = \frac{1}{d_i} \sum_{j:\, j \sim i} (x_j - \widehat{\mu}_i)(x_j - \widehat{\mu}_i)^\intercal. \tag{4}$$

3. The estimated local tangent direction, $\widehat{v}_i \in \mathbb{R}^d$, is the (normalized) eigenvector corresponding to the largest eigenvalue of $\widehat{\Sigma}_i$.

The vector $\widehat{v}_i$ represents the direction of maximum variance in the local neighborhood of x_i and thus serves as an approximation of the tangent vector to the underlying manifold $\mathcal{M}$ at or near x_i.

Remark 3.1 (Choice of neighborhood graph). In practice, r-neighborhood graphs and k-nearest-neighbor graphs perform similarly. However, the r-neighborhood graph has a symmetric adjacency matrix, which makes it easier to analyze. The hyperparameters k and r influence performance, much like the perplexity parameter in t-SNE—there is no universally optimal choice. We require k or r to be large enough to ensure the graph is connected. Figures 7 and 8 in the appendix indicate that neighborhood size (measured by k) is not crucial for our algorithm's performance. Under identical settings, our method consistently outperforms t-SNE.

3.2. Orientation Alignment

The local tangent vectors $\widehat{v}_i$ estimated in the previous step are unique only up to their sign, as $-\widehat{v}_i$ is also a valid principal eigenvector if $\widehat{v}_i$ is. To construct a meaningful global embedding, these local directions must be consistently oriented relative to each other across the manifold. Figure 1 (left) illustrates a set of such unaligned raw tangent vectors.

To achieve this global sign consistency, we seek a vector of signs $s = (s_1, \ldots, s_n)^\intercal$, where each $s_i \in \{-1, 1\}$, that maximizes the sum of dot products between signs of adjacent oriented tangents:

$$\max_{s \in \{-1,1\}^n} \sum_{i,j:\, i \sim j} s_i s_j \langle \widehat{v}_i, \widehat{v}_j \rangle. \tag{5}$$

This objective aims to make $s_i \widehat{v}_i$ and $s_j \widehat{v}_j$ point in similar directions if $\langle \widehat{v}_i, \widehat{v}_j \rangle > 0$, and opposite directions if

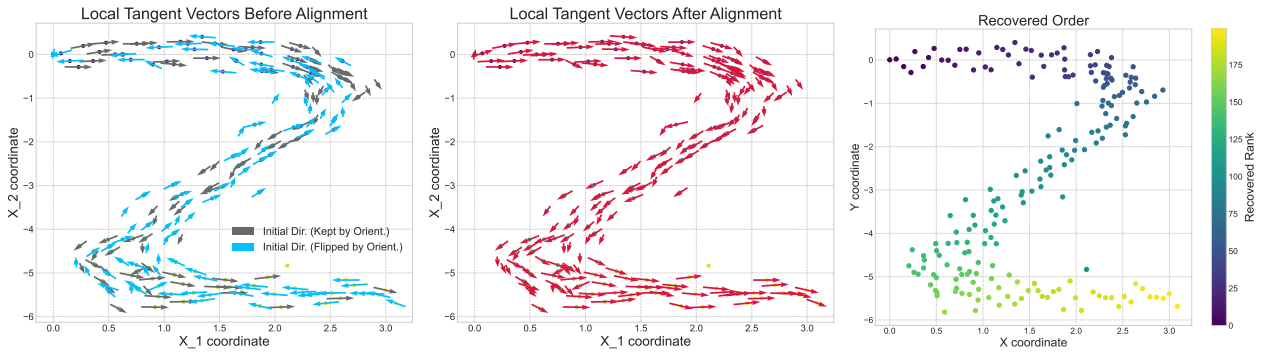


Figure 1. (Left) Initial PCA-derived tangent vectors. Their orientations are arbitrary before alignment. Colors indicate their fate after the orientation step: gray for vectors whose initial direction is kept, and blue for those that will be flipped. (Middle) Tangent vectors after spectral orientation. The vectors are now globally aligned along the manifold's intrinsic direction. (Right) Recovered order shown by color.

$\langle \widehat{v}_i, \widehat{v}_j \rangle < 0$. The problem in (5) is equivalent to maximizing $s^\intercal \widehat{Q} s$, where $\widehat{Q}$ is a symmetric affinity matrix with entries

$$\widehat{Q}_{ij} = \begin{cases} \langle \widehat{v}_i, \widehat{v}_j \rangle & \text{if } i \sim j \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Since optimizing over $s \in \{-1, 1\}^n$ is an NP-hard problem (related to Max-Cut), we employ a spectral relaxation. For any feasible s in (5), we have $\|s\|_2^2 = n$. Relaxing the constraint $s_i \in \{-1, 1\}$ to $\|s\|_2^2 = n$ (i.e., s lies on the sphere of radius $\sqrt{n}$), the problem becomes:

$$\widehat{u} = \arg \max_{s \in \mathbb{R}^n : \|s\|_2^2 = n} s^\intercal \widehat{Q} s. \tag{7}$$

The solution $\widehat{u}$ to this relaxed problem is the principal eigenvector of $\widehat{Q}$ (the eigenvector corresponding to the largest eigenvalue). The optimal signs for the original problem are then approximated by:

$$\widehat{s}_i = \operatorname{sign}(\widehat{u}_i), \quad \text{for } i = 1, \ldots, n. \tag{8}$$

This procedure yields the consistently oriented tangent vectors $\widehat{s}_i \widehat{v}_i$. Figure 1 (right) shows these tangent vectors after the alignment process.

3.3. Global Embedding via Least Squares

Having obtained consistently oriented local tangent vectors $\widehat{s}_i \widehat{v}_i$, we now seek a global 1-dimensional embedding $y = (y_1, \ldots, y_n)^\intercal \in \mathbb{R}^n$. The goal is that for any pair of neighboring points (x_i, x_j) (i.e., $i \sim j$), the difference $y_j - y_i$ in the embedding space should approximate the estimated "signed distance" or progression along the manifold from x_i to x_j.

A local estimate for this signed progression can be derived by projecting the displacement vector $x_j - x_i$ onto the local tangent at x_i, giving $\widehat{s}_i \widehat{v}_i^\intercal (x_j - x_i)$. Similarly, from x_j's perspective, it would be $\widehat{s}_j \widehat{v}_j^\intercal (x_j - x_i)$. To create a symmetric and robust local target for $y_j - y_i$, we average the two oriented tangent vectors and project the displacement onto this average direction. This defines our target local increment $\widehat{W}_{ij}$:

$$\widehat{W}_{ij} := \frac{1}{2} (\widehat{s}_i \widehat{v}_i + \widehat{s}_j \widehat{v}_j)^\intercal (x_j - x_i), \quad \text{for } i \sim j. \tag{9}$$

Note that $\widehat{W}_{ij} = -\widehat{W}_{ji}$ if $i \sim j$ is skew-symmetric. We then seek an embedding y by minimizing the sum of squared discrepancies between the embedding differences $y_j - y_i$ and these target increments $\widehat{W}_{ij}$ over all neighboring pairs:

$$\min_{y \in \mathbb{R}^n} \sum_{i,j:\, i \sim j} \left(y_j - y_i - \widehat{W}_{ij} \right)^2. \tag{10}$$

This objective function is invariant to a constant shift in y (i.e., if y is a solution, then $y + c\mathbf{1}$ is also a solution for any $c \in \mathbb{R}$), meaning the solution is not unique. To ensure a unique solution (up to global sign, which doesn't affect ordering), we impose a constraint, typically by centering the embedding: $\langle \mathbf{1}, y \rangle = \sum_i y_i = 0$. The optimization problem thus becomes:

$$\widehat{y} = \operatorname*{argmin}_{y \in \mathbb{R}^n :\, \langle \mathbf{1}, y \rangle = 0} \sum_{i,j:\, i \sim j} \left(y_j - y_i - \widehat{W}_{ij} \right)^2. \tag{11}$$

This is a standard least-squares problem whose solution can be found by solving a linear system related to the graph Laplacian. In implementation, one may solve either the direct least-squares problem in (11) or its Laplacian normal equations below; in our numerical experiments these two formulations produced identical orderings up to floating-point precision.

Lemma 3.2 (Laplacian form)**.** *Let A be the adjacency matrix of the neighborhood graph (where $A_{ij} = 1$ if $i \sim j$ and 0 otherwise), and let $L = D_A - A$ be the corresponding graph Laplacian matrix, where D_A is the diagonal matrix of node degrees. Let $\widehat{\eta} \in \mathbb{R}^n$ be a vector with entries*

$\widehat{\eta}_i = \sum_{j=1}^n A_{ij}\widehat{W}_{ij}$. *Minimizing the objective function in* (11) *is equivalent to*

$$\operatorname*{argmin}_{y\in\mathbb{R}^n:\ \langle \mathbf{1},y\rangle=0} \; y^T L y - 2y^T\widehat{\eta}. \tag{12}$$

The unique optimal solution $\widehat{y}$ *to* (11) *is given by* $\widehat{y} = L^+\widehat{\eta}$, *where* L^+ *is the Moore-Penrose pseudo-inverse of* L.

3.4. Final Ordering

Given the computed 1D embedding $\widehat{y} = (\widehat{y}_1, \dots, \widehat{y}_n)^\intercal$, the final estimated linear order of the original data points is obtained by sorting these embedding values. We define the permutation ϕ of the indices $\{1, \dots, n\}$ as:

$$\phi = \operatorname{argsort}_{j\in\{1,\dots,n\}} \widehat{y}_j. \tag{13}$$

This permutation ϕ provides the sequence $(\phi(1), \phi(2), \dots, \phi(n))$ such that $\widehat{y}_{\phi(1)} \le \widehat{y}_{\phi(2)} \le \dots \le \widehat{y}_{\phi(n)}$, representing the recovered seriation.

4. Theoretical Analysis

Consider $X = \gamma(\Theta) + \varepsilon$ with $\Theta \sim f_\Theta$ independent of the noise ε, of variance σ^2. Let $f_{\max}$ and $f_{\min}$ be the maximum and minimum of f_Θ, the density of Θ. We will make the following assumptions:

Assumption 4.1. γ is parameterized by arc-length, so that $\gamma'(\theta)$ is well-defined and $\|\gamma'(\theta)\| = 1$ for all θ. The curvature of γ is bounded by κ.

Assumption 4.2. The noise is bounded so that $\{x_i\}$ lie in a tubular neighborhood $\mathcal{T}_\rho = \{x \in \mathbb{R}^d :\ \operatorname{dist}(x, \mathcal{M}) \le \rho\}$ around the manifold, for some $\rho > 0$.

Assumption 4.3. For some $\delta \in (0, 1)$, and a sufficiently large numerical constant $C_1 > 0$, we have $(1-\delta)r \ge \rho + 2\sqrt{d}\sigma$ and $\sqrt{d}\sigma \le \delta r \le \frac{1}{\kappa}$ and

$$(\delta r)^3 \ge \frac{C_1}{f_{\min}} G_n(\sigma, \rho, \kappa) \quad \text{where}$$

$$G_n(\sigma, \rho, \kappa) := \sigma^2 + \rho^2 + \kappa^2(r^4 + \rho^4) + \sqrt{\frac{d}{n}}.$$

For r of constant order, these conditions are satisfied for sufficiently large n and small κ, ρ and σ. In particular, note that $G_n(\sigma, \rho, \kappa) \to 0$ as $n \to \infty$ and $\kappa, \sigma, \rho \to 0$.

We also need the following assumption to prevent the near self-intersection of the curve

Assumption 4.4. For any θ_0 in the domain of γ, the set $\{\theta :\ \gamma(\theta) \in B(\gamma(\theta_0), 2\rho + r)\}$ is included in the interval $[\theta_0 - \frac{\pi}{2\kappa}, \theta_0 + \frac{\pi}{2\kappa}]$.

This holds, for example, under the stronger assumption $2L\kappa \le \pi$ where L is the total length of the curve. This assumption implies that the curve does not come close to intersecting itself: For data points on γ that are close in $\mathbb{R}^d$, their inverse image (via γ^{-1}) are also close in the domain of γ. In other words, the neighborhood of any point on the curve large enough to capture noisy points contains only "one branch" of the curve.

We add a short interpretation to make the assumptions less opaque. Assumption 4.1 is a standard convention: any smooth curve with nonzero speed can be reparameterized by arc length, and the curvature bound κ then quantifies the geometric complexity entering the rate. Assumption 4.2 is used to ensure that local neighborhoods sample one tubular neighborhood of the curve, so local PCA estimates tangent directions rather than mixing unrelated parts of the curve. This type of tubular-noise model has precedent in manifold estimation, for example in the model class $\mathcal{P}^k(\sigma)$ of Aamari & Levrard (2019); related reach-based geometric conditions also appear throughout manifold learning (Niyogi et al., 2008; Aamari & Levrard, 2019). Boundedness is mainly a proof simplification: sub-Gaussian noise lies in such a tube with high probability, with ρ of order $\sigma\sqrt{d}$ up to logarithmic factors in n, although carrying this through the synchronization analysis would add technical bookkeeping. Assumption 4.3 calibrates the neighborhood radius r, requiring enough neighbors for stable PCA while keeping neighborhoods short relative to the curvature scale. Finally, Assumption 4.4 rules out near self-intersections at the estimation scale $2\rho + r$. This condition is related in spirit to a reach condition, but it is not the same object: it controls how the curve folds back on itself at the scale used by the graph, and ensures that ambient-space neighbors belong to the same local branch. Without such a condition, any local method would face an intrinsic ambiguity.

Let $d_{\text{avg},2} = \sqrt{\frac{1}{n}\sum_i d_i^2}$ be an average degree, and $\lambda_2(L)$ the spectral gap of the Laplacian L of the r-neighborhood graph $G_r(\mathscr{X}_n)$.

For vectors x and y, let $\tau(x, y)$ denote their Kendall's τ coefficient, a measure of their rank similarity, and set $\bar{\tau}(x, y) := 1 - \tau(x, y)$ for its complement. We have the following result:

Theorem 1 (Main result)**.** *Assume* $\kappa(2\rho + r) \ll 1$. *Then, with probability at least* 0.99, *we have*

$$\bar{\tau}(\widehat{y}, \theta - \bar{\theta}) \;\lesssim\; n^{-1/2} + f_{\max}^{2/3}\Big(\frac{d_{avg,2}}{\lambda_2(L)}\Big)^{2/3}\Big(K_\kappa(2\rho + r) + \frac{G_n(\sigma, \rho, \kappa)\, r}{f_{\min}(\delta r)^3} + \rho\Big)^{2/3}.$$

where $K_\kappa(t) := \kappa t^2(1 + \frac{1}{4}\kappa t)$.

See Appendix B for the proof. The bound follows from Proposition B.5, which converts MSE between vectors $\theta, y \in \mathbb{R}^n$ into a Kendall tau error bound $1 - \tau(\theta, y)$. Bounding $f_{\max}$ (the maximum density of Θ) is essential: without

it, MSE could be small while $1-\tau$ remains large. Intuitively, bounding $f_{\max}$ prevents θ_i from concentrating too much in a small region, which would allow large ranking changes with only small ℓ_2 movements.

The term

$$\frac{G_n}{f_{\min}(\delta r)^3} = \frac{\sigma^2+\rho^2+\kappa^2(r^4+\rho^4)}{f_{\min}(\delta r)^3} + \frac{\sqrt{d/n}}{f_{\min}(\delta r)^3}$$

captures tangent vector estimation error, where the first term is bias (approximation error) and the second is roughly variance (estimation error)—see Lemmas B.3 and B.4 and (17). The term $K_\kappa(2\rho+r)$ bounds the gap between $\widehat{W}_{ij}$ and its ideal value $\theta_j-\theta_i$. Even with no noise and exact tangent vectors, curvature κ creates a discrepancy, which $K_\kappa(2\rho+r)$ upper bounds (Lemma B.8). Note this vanishes as $\kappa \to 0$.

4.1. Interpretation of the result

To interpret the result, it helps to think of $f_{\min}$ and $f_{\max}$ as constants (the minimum and maximum of the density function of Θ): the ideal case is the uniform distribution on $[0, L]$, where $f_{\min} = f_{\max} = 1/L$ and L is the length of the curve, a constant. We also take r to be of constant order. Then, the quantity $\frac{d_{\text{avg},2}}{\lambda_2(L)}$ is expected to be of constant order (cf. (Von Luxburg et al., 2014)), hence can be absorbed into the constants. So the quantity $f_{\max}^{2/3}\left(\frac{d_{\text{avg},2}}{\lambda_2(L)}\right)^{2/3} \lesssim 1$ and can be ignored.

The quality of the bound is then controlled by the curvature κ, noise level (σ, ρ) and the sample size n. Consider the easiest case where $\gamma(\cdot)$ is a straight line, so that $\kappa = 0$ and we have no noise ($\sigma = 0$, so we can take $\rho = 0$ as well). In this case $K_\kappa(2\rho + r) = 0$ and $G_n(\sigma, \rho, \kappa) = \sqrt{d/n}$, and the the bound simply reads

$$\bar{\tau}(\widehat{y}, \theta - \bar{\theta}) \lesssim \left(\frac{\sqrt{d/n}\cdot r}{f_{\min}(\delta r)^3}\right)^{2/3} + \frac{1}{\sqrt{n}} \lesssim \left(\frac{d}{n}\right)^{1/3} + \frac{1}{\sqrt{n}}$$

where the second inequality further absorbed the constants r, δ and $f_{\min}$ into $\lesssim$. This bound shows that in the noiseless linear case, we achieve consistency as long as $n \to \infty$ (and $d = o(n)$).

Fundamental information limit. In the presense of noise, consistency is not possible: that is, adding (isotropic) noise, it will not be possible to recover the order consistently, in the sense of driving $1-\tau$ to zero. This is a fundamental information limit of the problem, experienced by any method, not a deficiency of our method. To see this, just consider the case where $\gamma(\theta) = \theta$, a straight line in $\mathbb{R}$. Adding noise ε_i to $\theta_i \sim U[0, 1]$ swaps the order of a positive fraction of pairs of points $\{\theta_i + \varepsilon_i\}_{i=1}^n$ as $n \to \infty$. A positive curvature exacerbates the problem (a smaller amount noise can swap the order for higher curvature).

A consequence of the above discussion is that for the $\sigma > 0$ regime, all one can hope for a method is to have an error that can be made arbitrarily small in the limit of $n \to \infty$, by making κ and σ small enough. This is exactly what Theorem 1 shows for STAGE. Asymptotically (as $n \to \infty$), we have

$$\lim_{n\to\infty} \bar{\tau}(\widehat{y}, \theta - \bar{\theta}) \lesssim \left(K_\kappa(2\rho+\sigma^2)+\rho^2+\kappa^2(1+\rho^4)+\rho\right)^{2/3}$$

and the RHS can be made arbitrarily small by making κ, σ and ρ small enough.

5. Experiments

We evaluate the empirical performance of our proposed STAGE algorithm against several existing seriation methods. All experiments were conducted on a desktop computer equipped with an AMD Ryzen 9 7900X processor and 32 GB of RAM.

5.1. Synthetic Data

We first assess STAGE on synthetic datasets, comparing with established algorithms: the spectral method by Atkins et al. (1998), 1D UMAP (McInnes et al., 2018), 1D t-SNE (Van der Maaten & Hinton, 2008), and the method by Recanati et al. (2018). We also compare with SABRE (Issartel et al., 2024), a recent seriation method with minimax guarantees for noisy Robinson matrices; see Appendix E.4. We include some experiments on curves in low dimensional simulation data in Section E.1 in the appendix, but here we focus on its performance in higher dimensions. Performance is quantified using Kendall's τ rank correlation coefficient between the true and recovered orderings.

For these experiments we utilize curves based on random Fourier series, letting the i-th dimension of γ take the form

$$\gamma_i(t) = \sum_{j=1}^{J} a_{ij}\sin(2\pi jt) + b_{ij}\cos(2\pi jt)$$

where $a_{ij}, b_{ij} \sim \mathcal{N}(0, j^{-\alpha})$ for some choice of α and K. By increasing K or decreasing α we can increase the curvature of the resulting manifold, giving us a modular method for creating curves in any dimension. These curves as described are closed, i.e. they begin and end at the same point with $t = 0$ and $t = 1$. To account for this we take the path on only a quarter of its span, between $t = 0$ to $t = 0.25$, to give us an open curve that is more suitable for the problem we aim to solve. We plot the first 3 dimensions of an example curve in Figure 5 in the appendix. Lastly, we modify our algorithm so that neighborhoods are built through k-NN with $k = 50$ rather than through a ball, as finding sufficiently many points in a ball can be difficult in high dimension. We test our method on curves

Table 1. Comparison of Methods ($n = 2000$), Kendall's τ and std. err. values multiplied by 100.

Dim.	Method	$\sigma = 0.5$	$\sigma = 1$	$\sigma = 1.5$	$\sigma = 2$
50	Fiedler Vector	85.47 (4.53)	76.91 (3.53)	49.49 (11.23)	33.22 (14.77)
	1D UMAP	88.6 (0.67)	77.08 (1.89)	64.5 (3.41)	51.17 (3.5)
	1D t-SNE	90.51 (0.63)	81.03 (1.43)	71.53 (2.75)	61.51 (3.16)
	Recanati	90.05 (1.03)	80.1 (1.35)	71.6 (1.86)	**64.39 (2.35)**
	STAGE	**90.91 (0.66)**	**81.76 (1.47)**	**72.26 (2.36)**	64.04 (2.58)
100	Fiedler Vector	83.1 (12.9)	83.34 (1.55)	63.0 (13.29)	44.9 (11.3)
	1D UMAP	90.89 (0.45)	81.72 (1.18)	71.14 (2.07)	58.85 (4.81)
	1D t-SNE	92.93 (0.33)	85.66 (0.86)	78.28 (1.49)	70.05 (3.35)
	Recanati	92.77 (0.47)	84.51 (0.76)	77.9 (1.03)	72.07 (1.39)
	STAGE	**93.37 (0.29)**	**86.46 (0.85)**	**79.47 (1.39)**	**72.27 (1.89)**
200	Fiedler Vector	68.8 (22.04)	88.7 (0.37)	81.2 (1.1)	66.52 (4.1)
	1D UMAP	93.29 (0.33)	86.35 (0.46)	78.89 (0.62)	70.17 (1.03)
	1D t-SNE	95.01 (0.08)	89.94 (0.26)	84.98 (0.28)	77.47 (4.18)
	Recanati	95.17 (0.12)	87.84 (0.21)	83.41 (0.34)	78.88 (0.49)
	STAGE	**95.47 (0.09)**	**90.77 (0.24)**	**86.12 (0.28)**	**81.18 (0.56)**

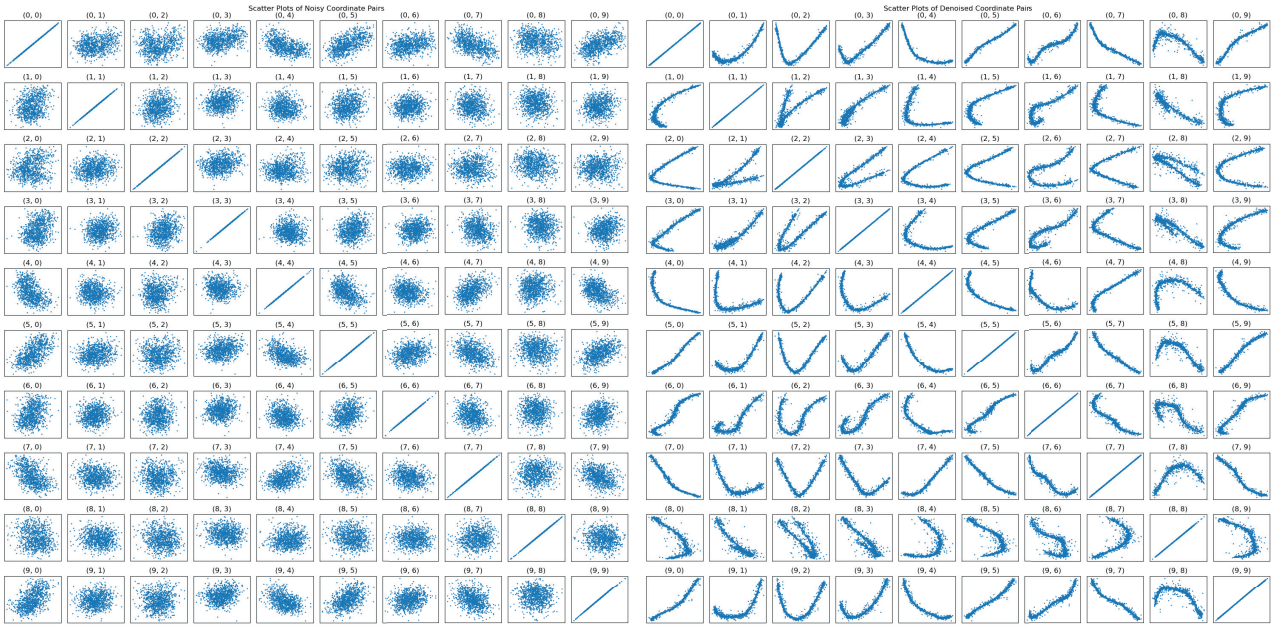
generated with $J = 10$ and $\alpha = 2.3$, and over dimensions $d \in \{50, 100, 200\}$, with noise levels $\sigma \in \{.5, 1, 1.5, 2\}$ and sample sizes $n \in \{500, 1000, 2000\}$, averaging over 10 runs with each combination of settings. In Table 1, we see that STAGE consistently beats the other methods in terms of τ for n = 2000 (the full results are displayed in Table 4 in the appendix with similar results), noting that STAGE takes significantly less time than the competitors to run, and that while our choice of neighborhood size is static, typically STAGE improves when neighborhood size increases with sample size, which we demonstrate experimentally in the appendix versus t-SNE. Figure 2 shows the results of applying STAGE to the first 10 dimensions of a random Fourier curve, demonstrating that STAGE not only recovers the correct ordering but also denoises the data. For a full comparison with the true curve, see Figure 6 in the appendix. We also performed an ablation in which the orientation-alignment step is removed before the final embedding. Across the same synthetic settings, this consistently led to lower Kendall's τ and substantially higher variance, confirming that the alignment stage is not merely a cosmetic preprocessing step but a key ingredient in the final recovery performance. Representative ablation results are reported in Table 8.

5.2. Real Data: Single-Cell RNA-Seq Pseudotime Ordering

To assess STAGE's applicability to real-world biological data, we applied it to the task of pseudotime ordering in single-cell RNA-Seq (scRNA-Seq) experiments. The goal is to infer a temporal progression of cells based on their gene expression profiles, reflecting processes such as development or differentiation. We utilized three different datasets to assess our method. The first is a publicly available dataset of human pre-implantation embryo development from Petropoulos et al. (2016), as processed and benchmarked in Saelens et al. (2019). After some additional processing and filtering, this dataset comprises 1289 cells and 6058 gene expression measurements, annotated with discrete developmental stages rather than a continuous ground-truth pseudotime. The other two datasets are subsets of the data collected by Packer et al. (2019), tracking the development of *C. elegans* cells over time. In particular, we consider the development of seam cells and hypodermis cells over time. The former contains 2766 cells and 14870 genes, and the latter 7746 cells and 16379 genes.

For each data set, we applied Principal Component Analysis (PCA) to the gene expression data for dimensionality reduction and visualization (Figure 3, left). We then ran STAGE on the first twenty principal components (PCs) to infer an ordering. We present the results on the three datasets in Table 2, where we can see that we perform slightly worse than t-SNE and Recanati on the Saelens dataset, but better on the two Packer datasets. We also include a visualization of the Saelens data in 2 PCs in Figure 3 which demonstrates that the STAGE embedding, when visualized against the known developmental stages (Figure 3, right), generally aligns with the expected biological progression. While some mixing of adjacent stages is observed, which is consistent with the inherent overlap and stochasticity in developmental processes and the discrete nature of the ground-truth labels, STAGE successfully captures the broad temporal trajectory.

An important consideration for such results is the selection of r. Briefly, we have found that it is important to select a radius such that the resulting neighborhood graph



(a) Noisy curve

(b) Denoised (STAGE)

Figure 2. Pairplots of the first 10 dimensions: (a) original noisy observations, (b) denoised points produced by STAGE.

is connected. In practice, a good radius may be somewhere between 1 and 2 times the minimum radius necessary to generate this connected neighborhood graph. We include further investigation of this in the appendix.

6. Discussion and Limitations

Our STAGE algorithm is tailored for linear seriation on 1D manifolds. While local PCA is applicable in higher dimensions, generalizing STAGE for k-dimensional manifold learning ($k > 1$) faces significant challenges.

Topological Constraint: Orientability. A key simplification for 1D manifolds is their inherent orientability, allowing for a globally consistent direction. Our spectral sign alignment (Section 3.2) leverages this to establish such a consistency. For non-orientable manifolds of dimension $k \geq 2$ (e.g., a 2D Möbius strip), a globally consistent orientation of tangent spaces does not exist. Thus, our current alignment strategy, which finds a single consistent orientation, would not directly apply. Determining manifold orientability from data is a non-trivial precursor task.

Methodological Generalization. Extending STAGE to $k > 1$ dimensions, even for orientable manifolds, presents further difficulties:

1. **Orientation Alignment:** Aligning k-dimensional tangent bases (frames) requires finding optimal orthogonal transformations (e.g., elements of $O(k)$ between neighbors. This is a more complex synchronization problem than the $\mathbb{Z}_2$ (sign) synchronization solved by Equation (7), demanding more sophisticated optimization techniques.

2. **Global Embedding:** The least-squares embedding (Section 3.3) would need to be reformulated to produce a k-dimensional output $y_i \in \mathbb{R}^k$, with the target increments $\widehat{W}_{ij}$ also becoming k-dimensional vectors.

Future Work Considerations. Future work might explore adapting STAGE for k-dimensional manifold learning, assuming orientability. This would involve developing methods for k-frame synchronization and a corresponding k-dimensional geometric embedding. However, the present work focuses on the specific advantages STAGE offers for the 1D linear seriation problem.

A related limitation concerns branching or more complex topologies. For a genuinely branching trajectory, such as a clean Y-shaped manifold, a single global total order may itself be ill-defined: points on different branches are often only partially ordered. In such cases, STAGE should be interpreted as producing a coarse one-dimensional linearization when a dominant backbone exists, rather than recovering the full branching structure. This is consistent with the intended scope of the method, which is linear seriation on a single latent 1D trajectory.

A natural theoretical question is whether one can obtain minimax rates for this point-cloud model. Existing minimax

	Fiedler	1D UMAP	1D t-SNE	Recanati	STAGE
Saelens	0.695	0.634	**0.808**	0.800	0.786
Packer Seam Cells	0.817	0.589	0.889	0.792	**0.926**
Packer Hypodermis	0.743	0.472	0.801	0.712	**0.822**

Table 2. Performance of seriation methods based on Kendall's τ on the three RNA-Seq datasets

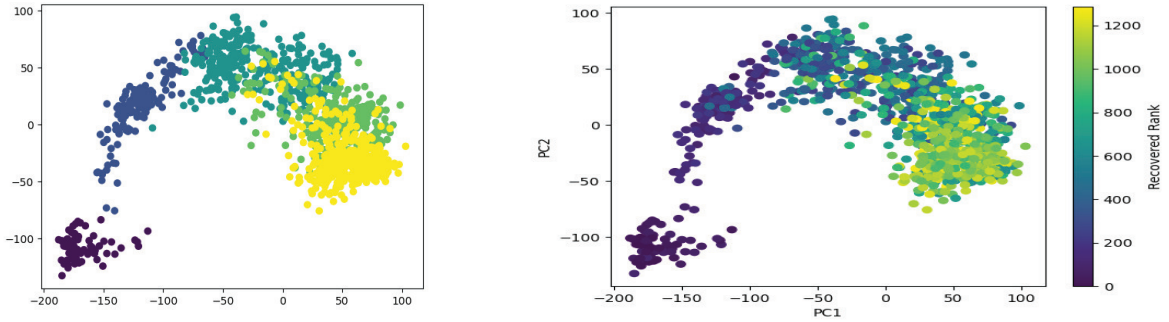


Figure 3. (Left) RNA-seq data reduced to 2 PCs and colored by actual stage, (Right) RNA-seq data reduced to 2 PCs and colored by recovered order

seriation results do not apply directly. Flammarion et al. (2019) observe a permuted noisy matrix $Y = \Pi^* A^* + Z$ and assume structural monotonicity or unimodality in the columns of A^*. If one tries to encode our model by taking $A_{ij} = \gamma_j(\theta_i)$, those assumptions generally fail: the coordinate functions of a curve in $\mathbb{R}^d$ can be multimodal even under the correct ordering, as in S-shaped curves, spirals, or the random Fourier curves used in our experiments. Issartel et al. (2024) study noisy Robinson matrices, where the similarity matrix is already assumed to decay away from the latent diagonal. In contrast, our observation model provides noisy point-cloud coordinates on an unknown embedded curve, rather than a noisy structured similarity matrix. A minimax analysis for this model remains open.

7. Conclusion

We presented STAGE (Spectral Tangent Alignment and Geometric Embedding), a geometric pipeline for linear seriation that works directly on noisy point clouds sampled from a 1D manifold. STAGE combines three stages: local tangent estimation via neighborhood PCA, spectral sign alignment for global orientation consistency, and a Laplacian least-squares embedding driven by oriented local projections. This preserves coordinate-level geometric cues that are often lost in similarity-matrix approaches, while remaining simple and efficient.

Our analysis provides finite-sample guarantees linking curvature, noise, neighborhood scale, and graph connectivity to rank recovery, yielding bounds on Kendall's τ that approach 1 under benign geometry and noise. Empirically, STAGE delivers accurate and fast order recovery on synthetic high-dimensional curves and real single-cell RNA-seq pseudotime data, often matching or surpassing strong baselines, and producing denoised embeddings that reflect manifold structure.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Award No. 1945667.

Impact Statement

This paper presents work whose goal is to advance the field of machine learning. There are many potential societal consequences of our work, none of which we feel must be specifically highlighted here.

References

Aamari, E. and Levrard, C. Nonasymptotic rates for manifold, tangent space and curvature estimation. *The Annals of Statistics*, 47(1):177 – 204, 2019.

Atkins, J. E., Boman, E. G., and Hendrickson, B. A spectral algorithm for seriation and the consecutive ones problem. *SIAM Journal on Computing*, 28(1):297–310, 1998.

Barnard, S. T., Pothen, A., and Simon, H. D. A spectral algorithm for envelope reduction of sparse matrices. In

Proceedings of the 1993 ACM/IEEE Conference on Supercomputing, pp. 493–502, 1993.

Belkin, M. and Niyogi, P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

Cai, T. T. and Ma, R. Matrix reordering for noisy disordered matrices: Optimality and computationally efficient algorithms. *IEEE Transactions on Information Theory*, 70(1): 509–531, 2023.

Clauset, A., Shalizi, C. R., and Newman, M. E. Power-law distributions in empirical data. *SIAM Review*, 51(4): 661–703, 2009.

Coifman, R. R. and Lafon, S. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006.

Coifman, R. R., Shkolnisky, Y., Sigworth, F. J., and Singer, A. Graph laplacian tomography from unknown random projections. *IEEE Transactions on Image Processing*, 17 (10):1891–1899, 2008.

Elson, J., Karp, R. M., Papadimitriou, C. H., and Shenker, S. Global synchronization in sensornets. In *Latin American Symposium on Theoretical Informatics*, pp. 609–624. Springer, 2004.

Flammarion, N., Mao, C., and Rigollet, P. Optimal rates of statistical seriation. *Bernoulli*, 25(1):623–653, 2019.

Fogel, F., Jenatton, R., Bach, F., and d'Aspremont, A. Convex relaxations for permutation problems. *Advances in Neural Information Processing Systems*, 26, 2013.

Garriga, G. C., Junttila, E., and Mannila, H. Banded structure in binary matrices. In *Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining*, pp. 292–300, 2008.

Giraud, C., Issartel, Y., and Verzelen, N. Localization in 1d non-parametric latent space models from pairwise affinities. *Electronic Journal of Statistics*, 17(1):1587–1662, 2023.

Issartel, Y., Giraud, C., and Verzelen, N. Minimax optimal seriation in polynomial time. *arXiv preprint arXiv:2405.08747*, 2024.

Janssen, J. and Smith, A. Reconstruction of line-embeddings of graphons. *Electronic Journal of Statistics*, 16(1):331–407, 2022.

Kruskal, J. B. and Wish, M. *Multidimensional scaling*. Number 07-011 in Sage University Paper Series on Quantitative Applications in the Social Sciences. Sage, 1978.

McInnes, L., Healy, J., and Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

Meidanis, J., Porto, O., and Telles, G. P. On the consecutive ones property. *Discrete Applied Mathematics*, 88(1-3): 325–354, 1998.

Natik, A. and Smith, A. Consistency of spectral seriation. *arXiv preprint arXiv:2112.04408*, 2021.

Niyogi, P., Smale, S., and Weinberger, S. Finding the homology of submanifolds with high confidence from random samples. *Discrete & Computational Geometry*, 39(1): 419–441, 2008.

Packer, J. S., Zhu, Q., Huynh, C., Sivaramakrishnan, P., Preston, E., Dueck, H., Stefanik, D., Tan, K., Trapnell, C., Kim, J., et al. A lineage-resolved molecular atlas of c. elegans embryogenesis at single-cell resolution. *Science*, 365(6459):eaax1971, 2019.

Petropoulos, S., Edsgärd, D., Reinius, B., Deng, Q., Panula, S. P., Codeluppi, S., Reyes, A. P., Linnarsson, S., Sandberg, R., and Lanner, F. Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell*, 165(4):1012–1026, 2016.

Recanati, A., Brüls, T., and d'Aspremont, A. A spectral algorithm for fast de novo layout of uncorrected long nanopore reads. *Bioinformatics*, 33(20):3188–3194, 2017.

Recanati, A., Kerdreux, T., and d'Aspremont, A. Reconstructing latent orderings by spectral clustering. *arXiv preprint arXiv:1807.07122*, 2018.

Robinson, W. S. A method for chronologically ordering archaeological deposits. *American Antiquity*, 16(4):293–301, 1951.

Roweis, S. T. and Saul, L. K. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290 (5500):2323–2326, 2000.

Saelens, W., Cannoodt, R., Todorov, H., and Saeys, Y. A comparison of single-cell trajectory inference methods. *Nature Biotechnology*, 37(5):547–554, 2019.

Salhov, M., Wolf, G., and Averbuch, A. Patch-to-tensor embedding. *Applied and Computational Harmonic Analysis*, 33(2):182–203, 2012. doi: 10.1016/j.acha.2011.11.003.

Schölkopf, B., Smola, A., and Müller, K.-R. Kernel principal component analysis. In *International conference on artificial neural networks*, pp. 583–588. Springer, 1997.

Singer, A. From graph to manifold laplacian: The convergence rate. *Applied and Computational Harmonic Analysis*, 21(1):128–134, 2006.

Singer, A. and Wu, H.-T. Vector diffusion maps and the connection laplacian. *Communications on Pure and Applied Mathematics*, 65(8):1067–1144, 2012.

Van der Maaten, L. and Hinton, G. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(11), 2008.

Vershynin, R. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge University Press, 2018.

Von Luxburg, U. A tutorial on spectral clustering. *Statistics and Computing*, 17:395–416, 2007.

Von Luxburg, U., Radl, A., and Hein, M. Hitting and commute times in large random neighborhood graphs. *Journal of Machine Learning Research*, 15(1):1751–1798, 2014.

Wolf, G. and Averbuch, A. Linear-projection diffusion on smooth euclidean submanifolds. *Applied and Computational Harmonic Analysis*, 34(1):1–14, 2013.

Zhang, Z. and Zha, H. Principal manifolds and nonlinear dimension reduction via local tangent space alignment. *SIAM Journal on Scientific Computing*, 26(1):313–338, 2004.

Álvarez Vizoso, J., Arn, R., Kirby, M., Peterson, C., and Draper, B. Geometry of curves in rn from the local singular value decomposition. *Linear Algebra and its Applications*, 571:180–202, 2019. ISSN 0024-3795.

A. On information loss in similarity-based approaches

Here we clarify in what sense working solely with a global similarity (or distance) matrix can discard geometric information that our method exploits, and illustrate this with a synthetic experiment.

Using the notation of Section 3.3, we note that $\widehat{W}_{ij}$ is the *signed distance along a tangent line*, and ideally given by $v_{ij}^T(x_i - x_j)$ where v_{ij} is a good estimate of the tangent direction near x_i and x_j. Our approach then solves

$$\widehat{y} = \arg \min_{y \in \mathbb{R}^n:\ \langle \mathbf{1}, y \rangle = 0} \sum_{i,j:\ i \sim j} \left(y_j - y_i - \widehat{W}_{ij} \right)^2.$$

On the other hand, a method that the relies on, say, a distance-based similarity (such as the Gaussian kernel), ideally tries to reconstructs a 1-D representation of the *squared distance*:

$$\widehat{y} = \arg \min_{y \in \mathbb{R}^n:\ \langle \mathbf{1}, y \rangle = 0} \sum_{i,j:\ i \sim j} \left((y_j - y_i)^2 - \|x_i - x_j\|^2 \right)^2.$$

Consider the case where x_i lie on a line and noise is orthogonal to the line: for example, $x_i = (\theta_i, 0) + (0, \varepsilon_i)$. Assuming we obtain a good estimate of v_{ij} (correctly aligned), we get $\widehat{W}_{ij} \approx \theta_j - \theta_i$ since v_{ij} nearly eliminates the noise. On the other hand, distance-based approaches observe $\|x_i - x_j\|^2 = (\theta_j - \theta_i)^2 + (\varepsilon_j - \varepsilon_i)^2$. For sufficiently large noise, the signal can be completely washed out.

Even if one can reconstruct the distances along the tangent line from the similarity matrix (e.g., (Recanati et al., 2017) attempts this by first embedding the points via Laplacian embedding), they would still optimize something like:

$$\widehat{y} = \arg \min_{y \in \mathbb{R}^n:\ \langle \mathbf{1}, y \rangle = 0} \sum_{i,j:\ i \sim j} \left((y_j - y_i)^2 - \widehat{W}_{ij}^2 \right)^2.$$

In general, $W_{ij} = \theta_i - \theta_j + \delta_{ij}$. Assuming δ_{ij} is nearly zero-mean, we have $\mathbb{E}[\widehat{W}_{ij}^2] \approx (\theta_i - \theta_j)^2 + \mathbb{E}[\delta_{ij}^2]$ for this approach versus $\mathbb{E}[\widehat{W}_{ij}] \approx (\theta_i - \theta_j)$ for our approach since $\mathbb{E}[\delta_{ij}] \approx 0$. Thus the squared distance approach suffers from the extra noise term $\mathbb{E}[\delta_{ij}^2]$.

It is worth noting that if the similarity matrix is already built from squared distances, the information is already lost in the anisotropic example and embedding back will not help.

Synthetic experiment. We illustrate this gap quantitatively. Consider equispaced points θ_i on $[0, 1]$ and let $x_i = (\theta_i, 0) + (\varepsilon_{i,\|}, \varepsilon_{i,\perp})$, with $\varepsilon_{i,\|} \sim \mathcal{N}(0, \sigma_\|^2)$ and $\varepsilon_{i,\perp} \sim \mathcal{N}(0, \sigma_\perp^2)$ independent. We fix $\sigma_\| = 10^{-3}$ and vary $\sigma_\perp$ over a grid $\sigma_\perp \in \{0.10, 0.13, \ldots, 0.35\}$. For Recanati's method (Recanati et al., 2018), we construct a Gaussian similarity matrix with bandwidth $\sigma = (\sigma_\|^2 + \sigma_\perp^2)^{1/2}$, while STAGE operates directly on the coordinates via local tangent estimation as in the main text. Figure 4 reports the Kendall's τ between the recovered order and the ground-truth order, averaged over 5 repetitions.

We observe that Recanati's τ degrades faster than STAGE. Only when $\sigma_\perp$ is so large that the geometry is essentially unrecoverable for both methods do their performances converge.

B. Detailed Proof of Theorem 1

Preliminary results. We start with a couple of lemmas. The first establishes the concentration of local emprical covariance matrix:

Lemma B.1 (Concentration)**.** *On the event* $|\{j : j \sim i\}| = d_i$*, With probability at least* $1 - 2e^{-\delta}$,

$$\|\widehat{\Sigma}_i - \Sigma(x_i, h)\| \lesssim \sqrt{\frac{d + \delta}{d_i}}.$$

Proof. The result follows from Exercise 4.7.3 of (Vershynin, 2018). □

The next result is well-known from differential geometry:

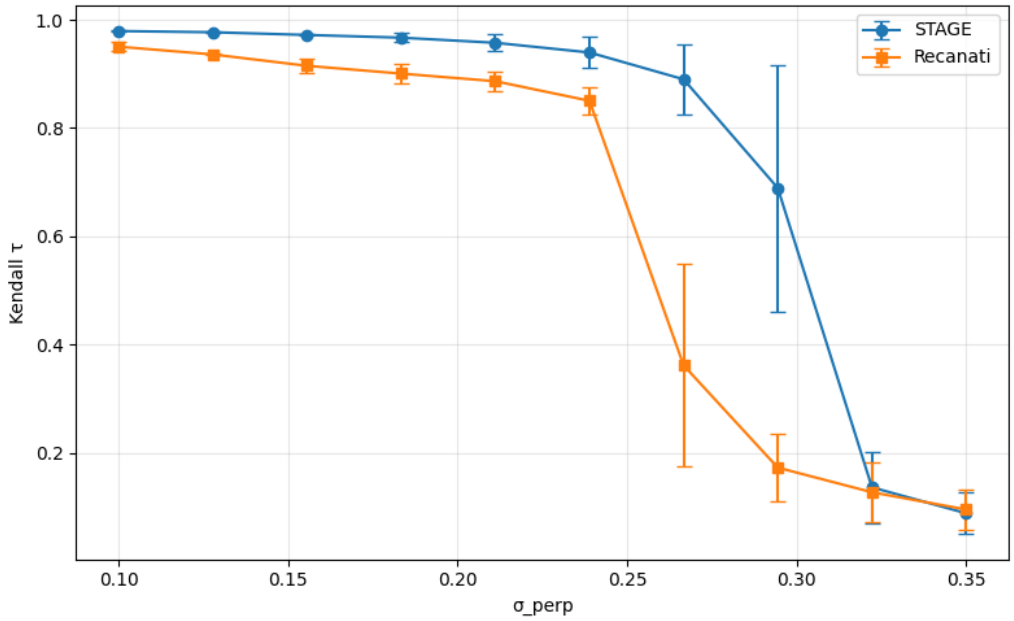


Figure 4. Kendall's τ between the recovered order and the ground-truth order, averaged over 5 repetitions, as a function of $\sigma_\perp$.

Lemma B.2 (Chord length inequality)**.** *Let $\gamma : I \to \mathbb{R}^d$ be a smooth curve parametrized by arc length s, defined on an interval I. Suppose the curvature $\kappa(s) = \|\gamma''(s)\|$ is bounded above by $\kappa > 0$. For any two parameter values $\theta_0, \theta \in I$ such that $|\theta - \theta_0| \le \pi/\kappa$, the following inequalities holds:*

$$\frac{2}{\kappa}\sin\Big(\frac{\kappa}{2}|\theta-\theta_0|\Big) \le \|\gamma(\theta)-\gamma(\theta_0)\|,$$
$$\frac{1}{\kappa}\sin(\kappa|\theta-\theta_0|) \le |\gamma'(\theta_0)^\intercal(\gamma(\theta)-\gamma(\theta_0))|.$$

We frequently combine the bounds with the elementary inequality $\sin(x) \ge \frac{2}{\pi}x$ for $0 \le x \le \pi/2$. In particular, assumuing $\kappa|\theta-\theta_0| \le \pi/2$, we have

$$\frac{2}{\pi}|\theta-\theta_0| \le \|\gamma(\theta)-\gamma(\theta_0)\| \tag{14}$$
$$\frac{2}{\pi}|\theta-\theta_0| \le |\gamma'(\theta_0)^\intercal(\gamma(\theta)-\gamma(\theta_0))|. \tag{15}$$

Let us now introduce some notation used in the spectral analysis of the local covariance matrices. For two (unit) vectors $u, v \in \mathbb{R}^d$, we write

$$d_\pm(u,v) := \min_{s\in\{\pm1\}} \|u - sv\|$$

Note that $|\langle u, v\rangle| = 1 - \frac{1}{2}d_\pm(u,v)$.

For any point $\theta \in [0,1]$, let $v(\theta) := \gamma'(\theta)$ be the tangent vector to the curve at θ.

For any point $x_0 \in \mathbb{R}^d$, other than the two endpoints, let $\pi(x_0)$ be a projection of x_0 onto the curve (a nearest point to x_0 on $\mathcal{M}$). Letting $\theta_0 = \gamma^{-1}(\pi(x_0))$, by smoothness of the curve, $v(\theta_0)$ will be orthogonal to $x_0 - \pi(x_0)$. Let $\rho_0 = \|x_0 - \pi(x_0)\|$ be the distance of x_0 to the curve.

We define

$$\Sigma(x_0, r) := \mathbb{E}[(x-x_0)(x-x_0)^T \mid x \in B(x_0, r)]$$

as the population version of (4), where $x = \gamma(\theta) + \varepsilon$. Here, x and θ are random variables, while x_0 and θ_0 are fixed.

Let $\sigma_1(x_0, r)$ be the top eigenvalue of $\Sigma(x_0, r)$. With x_0 and θ_0 related as described above, we let

$$\Sigma^*(x_0, r) = \sigma_1(x_0, r)\, v(\theta_0) v(\theta_0)^\intercal.$$

The next result provides a lower bound on the leading eigenvalue of $\Sigma(x_0, r)$, namely $\sigma_1(x_0, r)$.

Lemma B.3 (Top eigenvalue)**.** *Assume that* $(1-\delta)r \geq \rho + 2\sqrt{d}\sigma$ *for some* $\delta \in (0,1)$*. Then,*

$$\mathbb{E}[(\gamma'(\theta_0)^\intercal(x - x_0))^2 \mathbf{1}_{B_{x_0}(r)}(x)] \geq C_0\, f_{\min} \cdot u \cdot ((d\sigma^2) \vee u^2)$$

where $u = \frac{1}{\kappa} \wedge (\delta r)$ *for some numerical constant* $C_0 > 0$*. In particular, assuming* $\sqrt{d}\sigma \leq \delta r \leq \frac{1}{\kappa}$*, the lower boudns simplifies to* $C_0 f_{\min}(\delta r)^3$*.*

Proof. Recall that $\pi(x_0) = \gamma(\theta_0)$. Since $x_0 - \gamma(\theta_0)$ is perpendicular to $\gamma'(\theta_0)$, we have $\gamma'(\theta_0)^\intercal(x - x_0) = \gamma'(\theta_0)^\intercal(x - \gamma(\theta_0))$. Thus, we want to control

$$\mathbb{E}[(\gamma'(\theta_0)^\intercal(x - \gamma(\theta_0)))^2 \mathbf{1}_{B_{x_0}(r)}(x)] = \mathbb{E}[(\gamma'(\theta_0)^\intercal(\gamma(\theta) + \varepsilon - \gamma(\theta_0)))^2 \mathbf{1}_{B_{x_0}(r)}(x)].$$

Recall that θ and ε are random variables, while θ_0 and x_0 are fixed. Let us define the event

$$\mathcal{A} := \Big\{|\theta - \theta_0| \leq \min\Big(\frac{\pi}{2\kappa}, r - \rho - 2\sqrt{d}\sigma\Big)\Big\}, \quad \mathcal{E} := \big\{\|\varepsilon\| \leq 2\sqrt{d}\sigma\big\}.$$

By Lemma B.2, in particular (15), on $\mathcal{A}$, we have

$$|\gamma'(\theta_0)^\intercal(\gamma(\theta) - \gamma(\theta_0))| \geq \frac{2}{\pi}|\theta - \theta_0|.$$

Recall that $\gamma(\theta_0) = \pi(x_0)$, hence

$$x - x_0 = (\gamma(\theta) - \gamma(\theta_0)) + (\pi(x_0) - x_0) - \varepsilon. \tag{16}$$

We observe that

$$\|\varepsilon\| \leq 2\sqrt{d}\sigma \quad \text{and} \quad |\theta - \theta_0| \leq r - \rho - 2\sqrt{d}\sigma \implies \|x - x_0\| \leq r,$$

which follows from $\|\gamma(\theta) - \gamma(\theta_0)\| \leq |\theta - \theta_0|$ and $\|\pi(x_0) - x_0\| \leq \rho$ and the triangle inequality. This implies that $\mathcal{A} \cap \mathcal{E} \subseteq B_{x_0}(r)$.

Moreover, on $\mathcal{E}$, we have $|\gamma'(\theta_0)^\intercal \varepsilon| \leq 2\sqrt{d}\sigma$. Thus, on $\mathcal{A} \cap \mathcal{E}$, we have, by triangle inequality,

$$|\gamma'(\theta_0)^\intercal(\gamma(\theta) + \varepsilon - \gamma(\theta_0))| \geq \Big|\frac{2}{\pi}|\theta - \theta_0| - 2\sqrt{d}\sigma\Big|.$$

Squaring both sides and noting that $\mathbf{1}_{\mathcal{A}\cap\mathcal{E}} \leq \mathbf{1}_{B_{x_0}(r)}(x)$, we can derive the lower bound

$$\begin{aligned} F := \mathbb{E}[(\gamma'(\theta_0)^\intercal(\gamma(\theta) + \varepsilon - \gamma(\theta_0)))^2 \mathbf{1}_{B_{x_0}(r)}(x)] &\geq \mathbb{E}\Big[\Big(\frac{2}{\pi}|\theta - \theta_0| - 2\sqrt{d}\sigma\Big)^2 \mathbf{1}_{\mathcal{A}\cap\mathcal{E}}\Big]. \\ &= \mathbb{E}\Big[\Big(\frac{2}{\pi}|\theta - \theta_0| - 2\sqrt{d}\sigma\Big)^2 1_{\mathcal{A}}\Big]\, \mathbb{P}(\mathcal{E}) \end{aligned}$$

$\mathbb{P}(\mathcal{E})$ is lower bounded by an absolute constant. Let $b := \min\{\frac{\pi}{2\kappa}, r - \rho - 2\sqrt{d}\sigma\}$, so that $\mathcal{A} = \{|\theta - \theta_0| \leq b\}$. Then,

$$\begin{aligned} F_1 := \mathbb{E}\Big[\Big(\frac{2}{\pi}|\theta - \theta_0| - 2\sqrt{d}\sigma\Big)^2 1_{\mathcal{A}}\Big] &= \int f_\Theta(\theta)\Big(\frac{2}{\pi}|\theta - \theta_0| - 2\sqrt{d}\sigma\Big)^2 1_{\{|\theta-\theta_0|\leq b\}} d\theta \\ &\geq 2 f_{\min} \int_0^b \Big(\frac{2}{\pi}t - 2\sqrt{d}\sigma\Big)^2 dt \\ &\geq f_{\min}\frac{\pi}{3}\left(\Big(\frac{2}{\pi}b - 2\sqrt{d}\sigma\Big)^3 + (2\sqrt{d}\sigma)^3\right). \end{aligned}$$

Let $\frac{2}{\pi}b = x$ and $2\sqrt{d}\sigma = a$. We have $(x-a)^3 + a^3 = xh(x)$ where $h(x)$ is the quadratic function $x^2 - 3ax + 3a^2$, which has a global minimum of $\frac{3}{4}a^2$. We also note that $h(x) - \frac{x^2}{4} = \frac{3}{4}(x-2a)^2 \geq 0$. Thus, we have $h(x) \geq \frac{1}{4}\max\{3a^2, x^2\}$.

By assumption $(1-\delta)r \geq \rho + 2\sqrt{d}\sigma$, we have $b \geq \min\{\frac{\pi}{2\kappa}, \delta r\} \geq \frac{1}{\kappa} \wedge (\delta r) =: u$. We obtain

$$\begin{aligned} F_1 &\geq f_{\min}\frac{\pi}{3}x \cdot \frac{1}{4}\max\{3a^2, x^2\} \\ &= f_{\min} \cdot \frac{b}{6}\max\{3(2\sqrt{d}\sigma)^2, (2b/\pi)^2\} \gtrsim f_{\min}u((d\sigma^2) \vee u^2). \end{aligned}$$

Combined with $\mathbb{P}(\mathcal{E}) \gtrsim 1$, the proof is complete. □

The next lemma provides an upper bound on the *second* largest eigenvalue of $\Sigma(x_0, r)$:

Lemma B.4 (Second largest eigenvalue)**.** *Under Assumption 4.4, for any v perpendicular to $\gamma'(\theta_0)$,*

$$\mathbb{E}[(v^\intercal(x-x_0))^2\mathbf{1}_{B_{x_0}(r)}(x)] \leq C_0'\Big(\sigma^2 + \rho^2 + \kappa^2(r^4+\rho^4)\Big)$$

for some numerical constant $C_0' > 0$.

Proof. Given the above setting, we want to find an upper bound for

$$\mathbb{E}[(v^\intercal(x-x_0))^2\mathbf{1}_{B_{x_0}(r)}(x)] = \mathbb{E}[(v^\intercal((x_0-\gamma(\theta_0)) + (\gamma(\theta_0)-\gamma(\theta)) + \varepsilon))^2\mathbf{1}_{B_{x_0}(r)}(x)].$$

for arbitrary unit v perpendicular to $\gamma(\theta_0)$. We can use the inequality $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$, so we analyze term by term. We have

$$\mathbb{E}[(v^\intercal(x_0-\gamma(\theta_0))^2\mathbf{1}_{B_{x_0}(r)}(x)] \leq \rho^2$$

and

$$\mathbb{E}[(v^\intercal\varepsilon)^2\mathbf{1}_{B_{x_0}(r)}(x)] \leq \sigma^2.$$

It remains to consider the term $\mathbb{E}[(v^\intercal(\gamma(\theta)-\gamma(\theta_0))^2\mathbf{1}_{B_{x_0}(r)}(x)]$.

From (16), we note that $\|x_0 - x\| \leq r$ implies

$$\|\gamma(\theta)-\gamma(\theta_0)\| \leq r + \rho + \|\varepsilon\| \leq r + 2\rho$$

using $\|\varepsilon\| \leq \rho$ and $\|\pi(x_0) - x_0\| \leq \rho$. Assumption 4.4 then implies $|\theta - \theta_0| \leq \frac{\pi}{2\kappa}$.

Using a second order Taylor expansion of $\gamma(\theta)$ around θ_0, and the fact that $v^\intercal\gamma'(\theta_0) = 0$ and $\|\gamma''(\cdot)\| \leq \kappa$, we have

$$\begin{aligned} \mathbb{E}[(v^\intercal(\gamma(\theta)-\gamma(\theta_0))^2\mathbf{1}_{B_{x_0}(r)}(x)] &\leq \frac{\kappa^2}{4}\mathbb{E}\left[(\theta-\theta_0)^4\mathbf{1}_{B_{x_0}(r)}(x)\right] \\ &\leq \frac{\kappa^2}{4}\frac{\pi^4}{16}\mathbb{E}\left[\|\gamma(\theta)-\gamma(\theta_0)\|^4\mathbf{1}_{B_{x_0}(r)}(x)\right] \end{aligned}$$

where the second inequaly uses Lemma B.2, in particular (14), and the fact that for $x \in B_{x_0}(r)$, the condition $|\theta - \theta_0| \leq \frac{\pi}{2\kappa}$ holds.

Using the fact that $\|\gamma(\theta)-\gamma(\theta_0)\|^4\mathbf{1}_{B_{x_0}(r)}(x) \leq (r+2\rho)^4 \lesssim (r^4+\rho^4)$ finishes the proof. □

The following proposition is key in converting the bound on MSE to the bound on Kendall's tau:

Proposition B.5 (From MSE to Kendall's tau)**.** *Let F be a distribution on a compact interval $[a,b] \subset \mathbb{R}$ with density f satisfying $f(x) \leq f_{\max} < \infty$. Let $\theta = (\theta_1, \ldots, \theta_n)$ where θ_i are i.i.d. draws from F. Consider another vector $y = (y_i) \in \mathbb{R}^n$ and let $\varepsilon_n^2 = \frac{1}{n}\sum_{i=1}^n (y_i - \theta_i)^2$. Then, with probability at least $1-\delta$,*

$$1 - \tau(y,\theta) \leq 16 \cdot 2^{-1/3}\big(\varepsilon_n f_{\max}\big)^{2/3} + C\sqrt{\log(1/\delta)}\, n^{-1/2}$$

where $C > 0$ is a universal constant.

Main argument. We are now ready to prove the theorem. Fix x_i. We apply Lemmas B.3 and B.4 with $x_0 = x_i$ and $\theta_0 = \gamma^{-1}(\pi(x_i)) =: \theta_i^*$. We also let $x_i^* := \pi(x_i)$ and $v_i^* = v(\theta_i^*)$.

By Assumption 4.3 and Lemma B.3, we have $\sigma(x_i, r) \geq C_0 f_{\min}(\delta r)^3$, and by Lemma B.4,

$$\|\Sigma(x_i, r) - \Sigma^*(x_i, r)\| \lesssim \sigma^2 + \rho^2 + \kappa^2(r^4 + \rho^4) =: D_1.$$

By Lemma B.1, and the fact that w.h.p. $d_i \gtrsim n$, we have w.h.p,

$$\|\widehat{\Sigma}_i - \Sigma^*(x_i, r)\| \lesssim D_1 + \sqrt{\frac{d}{n}} = G_n(\sigma, \rho, \kappa),$$

where $G_n(\sigma, \rho, \kappa)$ is defined in Assumption 4.3. Note that $\Sigma^*(x_i, r)$ is a rank-one matrix with leading eigenvector $v_i^* = v(\theta_i^*)$ and eigengap $\sigma_1(x_i, r) - 0$. Applying Davis–Kahan theorem, for some numerical constant $C_2 > 0$, we have

$$d_\pm(\widehat{v}_i, v_i^*) \lesssim \frac{\|\widehat{\Sigma}_i - \Sigma^*(x_i, r)\|}{\sigma_1(x_i, r)} \leq C_2 \frac{G_n(\sigma, \rho, \kappa)}{f_{\min}(\delta r)^3} \leq \frac{C_2}{C_1} < \frac{1}{4} \tag{17}$$

where the two final inequalities follow by applying Assumption 4.3 and taking C_1 large enough in that assumption.

Let $s_i^* \in \{\pm\}$ be the optimal sign between $\widehat{v}_i$ and v_i^*, that is,

$$\|s_i^* \widehat{v}_i - v_i^*\| = d_\pm(\widehat{v}_i, v_i^*).$$

A curvature argument gives:

Lemma B.6. $\langle v_j^*, v_i^* \rangle \geq \frac{1}{2}$ *for* $i \sim j$.

Combining (17) and the preceding lemma, and applications of triangle inequality, we have $\langle s_i^* \widehat{v}_i, s_j^* \widehat{v}_j \rangle > 0$ for all i, j. Now we have:

Lemma B.7 (Global sign consistency)**.** *Let* $\widehat{s} = \operatorname{sign}(\widehat{u})$ *where* $\widehat{u}$ *is the solution of* (7)*. Then,* $\widehat{s} = s^*$.

Proof. Consider the change of variable $\xi = (\xi_i)$ where $\xi_i = s_i^* s_i$ where $s = (s_i)$ is the optimization variable in (7). Note also that $s_i = s_i^* \xi_i$ since $s_i^* \in \{\pm 1\}$. This leads to the optimization $\widehat{\xi} = \operatorname{argmax}_{\|\xi\|^2 = n} \xi^T W \xi$ where $W_{ij} = s_i^* s_j^* \widehat{W}_{ij} = \langle s_i^* \widehat{v}_i, s_j^* \widehat{v}_j \rangle \cdot 1\{i \sim j\}$. By the above argument W is a nonnegative matrix. Recall that we assumed $G_r(\mathscr{X}_n)$ is regular, and in particular connected. Thus W is a irreducible. By Perron–Frobenius theorem, its leading eigenvector has elements that are all positive, that is, $\operatorname{sign}(\widehat{\xi}_i) = 1$ for all $i \in [n]$. Since $\widehat{u}_i = s_i^* \widehat{\xi}_i$, it follows that $\operatorname{sign}(\widehat{u}_i) = s_i^*$. The proof is complete. □

Thus, we now have

$$\|\widehat{s}_i \widehat{v}_i - v_i^*\| = d_\pm(\widehat{v}_i, v_i^*) \leq C_2 \frac{G_n(\sigma, \rho, \kappa)}{f_{\min}(\delta r)^3} \quad \forall i \in [n]. \tag{18}$$

Recall that $\theta_i^* := \gamma^{-1}(x_i^*)$ and $v_i^* = \gamma'(\theta_i^*)$. Recall also that the original $\theta_i = \gamma^{-1}(x_i)$. Let us define:

$$W_{ij}^* = \frac{1}{2}(v_i^* + v_j^*)^T (x_j^* - x_i^*) \cdot 1\{i \sim j\}. \tag{19}$$

Lemma B.8. *Let* $h_* = \max_{(i,j): i \sim j} |\theta_i^* - \theta_j^*|$*, and write* $K_\kappa(t) := \kappa t^2 (1 + \frac{1}{4}\kappa t)$*. Then,*

$$\max_{(i,j): i \sim j} |W_{ij}^* - (\theta_j^* - \theta_i^*)| \leq K_\kappa(h_*).$$

Recall that $\gamma(\theta_i^*) = \pi(x_i) =: x_i^*$ for all i. To control h_*, we first note that, for $i \sim j$,

$$\begin{aligned} \|\gamma(\theta_i^*) - \gamma(\theta_j^*)\| &= \|\pi(x_i) - \pi(x_j)\| \\ &\leq \|\pi(x_i) - x_i\| + \|x_i - x_j\| + \|x_j - \pi(x_j)\| \leq \rho + r + \rho. \end{aligned}$$

By Assumption 4.4, we then have $|\theta_i^* - \theta_j^*| \le \frac{\pi}{2\kappa}$. Hence, for $i \sim j$, we can apply the chord length inequality (Lemma B.2), namely (14), to obtain

$$|\theta_i^* - \theta_j^*| \le \frac{\pi}{2}\|\pi(x_i) - \pi(x_j)\| \lesssim 2\rho + r.$$

Hence, we can take $h_* \lesssim 2\rho + r$ in Lemma (B.8). By a similar argument, we have

$$\begin{aligned}\|\gamma(\theta_i^*) - \gamma(\theta_i)\| &\le \|\pi(x_i) - x_i\| + \|x_i - \gamma(\theta_i)\| \\ &\le \rho + \|\varepsilon_i\| \le 2\rho\end{aligned}$$

By Assumption 4.4, we then have $|\theta_i^* - \theta_i| \le \frac{\pi}{2\kappa}$. We can apply the chord length inequality (Lemma B.2), namely (14), to obtain

$$|\theta_i^* - \theta_i| \le \frac{\pi}{2}\|\gamma(\theta_i^*) - \gamma(\theta_i)\| \le \pi\rho,$$

hence $|(\theta_i^* - \theta_j^*) - (\theta_i - \theta_i)| \le 2\pi\rho \lesssim \rho$. Combining these results we have

$$\max_{(i,j):i\sim j} |W_{ij}^* - (\theta_j - \theta_i)| \lesssim K_\kappa(2\rho + r) + \rho.$$

Next, we note that, for $i \sim j$

$$\begin{aligned}2|\widehat{W}_{ij} - W_{ij}^*| &\le \left\|(\widehat{s}_i\widehat{v}_i) + \widehat{s}_j\widehat{v}_j - (v_i^* + v_j^*)\right\| \cdot \|x_j - x_i\| \\ &\qquad + \|v_i^* + v_j^*\| \cdot \left\|(x_j - x_i) + (x_j^* - x_i^*)\right\| \\ &\le 2C_2\frac{G_n(\sigma,\rho,\kappa)}{f_{\min}(\delta r)^3} \cdot r + 2 \cdot 2\rho\end{aligned}$$

using (18) twice for the first term (via triangle inequality), and $\|v_i^*\| = \|v_j^*\| = 1$ for the second term. Putting the pieces together

$$|\widehat{W}_{ij} - (\theta_j - \theta_i)| \lesssim K_\kappa(2\rho + r) + \frac{G_n(\sigma,\rho,\kappa) \cdot r}{f_{\min}(\delta r)^3} + \rho, \quad i \sim j. \tag{20}$$

Now, we can use the following perturbation result:

Lemma B.9. *Assume $\widehat{W}_{ij} = (\theta_j - \theta_i) + \Delta_{ij}$ for $i \sim j$ and let $\widehat{y}$ be the unique solution of (11). Then, $\widehat{y} = \theta - \bar{\theta} + L^+\eta$ where $\eta_i = \sum_j A_{ij}\Delta_{ij}$. Moreover, with probability at least* 0.99*,*

$$1 - \tau(\widehat{y}, \theta - \bar{\theta}) \le 16 f_{\max}^{2/3}\left(\frac{d_{avg,2}}{\lambda_2(L)}\right)^{2/3}\|\Delta\|_\infty^{2/3} + C\sqrt{\log(100)}\, n^{-1/2}$$

where $\|\Delta\|_\infty = \max_{i,j}|\Delta_{ij}|$ and $d_{avg,2} = \sqrt{\frac{1}{n}\sum_i d_i^2}$ is an average degree of the graph.

Applying this result with $\|\Delta\|_\infty \lesssim$ RHS of (20), finishes the proof.

C. Remaining Proofs

C.1. Proof of Lemma 3.2

Let $J(y) = \sum_{i,j:\, i\sim j}\left(y_j - y_i - \widehat{W}_{ij}\right)^2$ be the objective function in (11). Since $i \sim j$ implies $A_{ij} = 1$, we can write this sum over all pairs (i,j) using A_{ij}:

$$\begin{aligned}J(y) &= \sum_{i=1}^n\sum_{j=1}^n A_{ij}\left(y_j - y_i - \widehat{W}_{ij}\right)^2 \\ &= \sum_{i,j} A_{ij}\left((y_j - y_i)^2 - 2(y_j - y_i)\widehat{W}_{ij} + \widehat{W}_{ij}^2\right) \\ &= \sum_{i,j} A_{ij}(y_j - y_i)^2 - 2\sum_{i,j} A_{ij}(y_j - y_i)\widehat{W}_{ij} + \sum_{i,j} A_{ij}\widehat{W}_{ij}^2.\end{aligned}$$

The first term is a standard quadratic form for the graph Laplacian: $\sum_{i,j} A_{ij}(y_j - y_i)^2 = 2y^T Ly$. (This follows from $\sum_{i,j} A_{ij}(y_j^2 - 2y_iy_j + y_i^2) = \sum_j y_j^2(\sum_i A_{ij}) - 2\sum_{i,j} A_{ij}y_iy_j + \sum_i y_i^2(\sum_j A_{ij}) = \sum_j d_jy_j^2 - 2y^T Ay + \sum_i d_iy_i^2 = 2y^T D_A y - 2y^T Ay = 2y^T(D_A - A)y = 2y^T Ly$.)

For the second term (the cross term):

$$-2\sum_{i,j} A_{ij}(y_j - y_i)\widehat{W}_{ij} = -2\sum_{i,j} A_{ij}y_j\widehat{W}_{ij} + 2\sum_{i,j} A_{ij}y_i\widehat{W}_{ij}.$$

The second part of this expression is $2\sum_i y_i \left(\sum_j A_{ij}\widehat{W}_{ij}\right) = 2\sum_i y_i\widehat{\eta}_i = 2y^T\widehat{\eta}$. For the first part, $-2\sum_{i,j} A_{ij}y_j\widehat{W}_{ij}$, we swap indices $i \leftrightarrow j$: $-2\sum_{j,i} A_{ji}y_i\widehat{W}_{ji}$. Since $A_{ji} = A_{ij}$ (undirected graph) and $\widehat{W}_{ji} = -\widehat{W}_{ij}$ (as $\widehat{W}_{ij} = \frac{1}{2}(\widehat{s}_i\widehat{v}_i + \widehat{s}_j\widehat{v}_j)^\intercal(x_j - x_i)$), this becomes: $-2\sum_{i,j} A_{ij}y_i(-\widehat{W}_{ij}) = 2\sum_{i,j} A_{ij}y_i\widehat{W}_{ij} = 2\sum_i y_i \left(\sum_j A_{ij}\widehat{W}_{ij}\right) = 2y^T\widehat{\eta}$. Thus, the entire cross term is $2y^T\widehat{\eta} + 2y^T\widehat{\eta} = 4y^T\widehat{\eta}$.

The third term, $\sum_{i,j} A_{ij}\widehat{W}_{ij}^2$, is a constant with respect to y. Therefore, $J(y) = 2y^T Ly - 4y^T\widehat{\eta} + \text{constant}$. Minimizing $J(y)$ subject to $\langle \mathbf{1}, y\rangle = 0$ is equivalent to minimizing $2y^T Ly - 4y^T\widehat{\eta}$ subject to the same constraint, which is further equivalent to minimizing $y^T Ly - 2y^T\widehat{\eta}$ subject to $\langle \mathbf{1}, y\rangle = 0$. This establishes the equivalence stated in (12).

Now, we find the optimal solution. We want to minimize $f(y) = y^T Ly - 2y^T\widehat{\eta}$ subject to $\mathbf{1}^T y = 0$. We form the Lagrangian: $\mathcal{L}(y, \lambda) = y^T Ly - 2y^T\widehat{\eta} - \lambda(\mathbf{1}^T y)$. Taking the derivative with respect to y and setting to zero: $\nabla_y\mathcal{L} = 2Ly - 2\widehat{\eta} - \lambda\mathbf{1} = 0 \implies Ly = \widehat{\eta} + \frac{\lambda}{2}\mathbf{1}$. Let $\lambda' = \lambda/2$. So, $Ly = \widehat{\eta} + \lambda'\mathbf{1}$. For this system to have a solution, the right-hand side must be orthogonal to the null space of L. Assuming the graph is connected, the null space of L is spanned by $\mathbf{1}$. Thus, $\mathbf{1}^T(\widehat{\eta} + \lambda'\mathbf{1}) = 0 \implies \mathbf{1}^T\widehat{\eta} + \lambda'\mathbf{1}^T\mathbf{1} = 0 \implies \mathbf{1}^T\widehat{\eta} + n\lambda' = 0$.

We show that $\mathbf{1}^T\widehat{\eta} = 0$: $\mathbf{1}^T\widehat{\eta} = \sum_i \widehat{\eta}_i = \sum_i\sum_j A_{ij}\widehat{W}_{ij}$. Let $M_{ij} = A_{ij}\widehat{W}_{ij}$. Then $M_{ji} = A_{ji}\widehat{W}_{ji} = A_{ij}(-\widehat{W}_{ij}) = -M_{ij}$. The matrix M is anti-symmetric. The sum of all entries of an anti-symmetric matrix is zero, i.e., $\sum_i\sum_j M_{ij} = 0$. Thus, $\mathbf{1}^T\widehat{\eta} = 0$.

Since $\mathbf{1}^T\widehat{\eta} = 0$, the condition $n\lambda' = 0$ implies $\lambda' = 0$ (assuming $n > 0$). The system simplifies to $Ly = \widehat{\eta}$. Since $\mathbf{1}^T\widehat{\eta} = 0$, $\widehat{\eta}$ is in the column space of L. The general solution to $Ly = \widehat{\eta}$ is $y = L^+\widehat{\eta} + c\mathbf{1}$ for some constant $c \in \mathbb{R}$, where L^+ is the Moore-Penrose pseudo-inverse of L. Applying the constraint $\mathbf{1}^T y = 0$: $\mathbf{1}^T(L^+\widehat{\eta} + c\mathbf{1}) = 0 \implies \mathbf{1}^T L^+\widehat{\eta} + c\mathbf{1}^T\mathbf{1} = 0$. Since $L^+\mathbf{1} = \mathbf{0}$ (a property of the pseudo-inverse of a graph Laplacian), we have $\mathbf{1}^T L^+ = (L^+\mathbf{1})^T = \mathbf{0}^T$. So, $\mathbf{0}^T\widehat{\eta} + cn = 0 \implies cn = 0 \implies c = 0$. Therefore, the unique optimal solution satisfying the constraint is $\widehat{y} = L^+\widehat{\eta}$.

C.2. Proof of Proposition B.5

For any $t > 0$, let $I_{ij}(t) = 1\{|\theta_i - \theta_j| \le 2t\}$ For $i \ne j$, by independence of θ_i and θ_j,

$$\mathbb{E}\big[I_{ij}(t)\big] \;=\; \iint_{|x-y|\le 2t} f(x)f(y)\,dx\,dy \;=\; \int f(x)\Big[\int_{x-2t}^{x+2t} f(y)\,dy\Big]dx \;\le\; 4t\,f_{\max}.$$

Let $U(t) = \frac{1}{\binom{n}{2}}\sum_{i<j} I_{ij}(t)$ which is an order statistic of order 2, with $\mathbb{E}[U(t)] \le 4t\,f_{\max}$. By McDiarmid's (a.k.a. bounded difference) inequality ,

$$\mathbb{P}\big[U(t) \ge 4tf_{\max} + u\big] \le \exp\big(-c\,nu^2\big).$$

Consider the set $\Lambda(t) = \{i < j : |\theta_i - \theta_j| \le 2t\}$. Then, $|\Lambda(t)| = \binom{n}{2}U(t)$. taking $u = \sqrt{\frac{\log(1/\delta)}{cn}}$, we have with probability at least $1 - \delta$,

$$|\Lambda(t)| \le \binom{n}{2}\Big(4tf_{\max} + \sqrt{\frac{\log(1/\delta)}{cn}}\Big) \le 2tf_{\max}n^2 + C_1\sqrt{\log(1/\delta)}\,n^{3/2}$$

for some constant C_1.

Now, let $B(t) = \{i : |y_i - \theta_i| \ge t\}$. Then, $|B(t)| \le n\varepsilon_n^2/t^2$ by Markov inequality (applied to the discrete measure $(1/n, \ldots, 1/n)$).

Let D be the set of discordant pairs, that is, $D = \{i < j :\ (y_i - y_j)(\theta_i - \theta_j) < 0\}$. A pair (i, j) is discordant only if

- Set D_1: either endpoint is in $B(t)$ or
- Set D_2: both endpoints are in $[n] \setminus B(t)$ and $|\theta_i - \theta_j| \le 2t$.

that is, $D \subset D_1 \cup D_2$. To see this let $d^\theta_{ij} = \theta_i - \theta_j$ and $d^y_{ij} = y_i - y_j$ and $e_i = y_i - \theta_i$. Then, $d^y_{ij} = d^\theta_{ij} + e_i - e_j$. Suppose that (i,j) is in neither D_1 nor D_2. Then, $|e_i| < t$ and $|e_j| < t$ and $|d^\theta_{ij}| > 2t$. Since then $|d^y_{ij} - d^\theta_{ij}| < 2t$, d^y_{ij} has the same sign as d^θ_{ij}. Thus, $D_1^c \cap D_2^c \subset D^c$ which is equivalent to the claim.

We have $D_1 \subset \{i < j : i \in B(t) \text{ or } j \in B(t)\}$, hence $|D_1| \le nB(t)$. Also, $D_2 \subset \Lambda(t)$, hence

$$|D| \le n|B(t)| + |\Lambda(t)|$$

and for $n \ge 2$,

$$1 - \tau(y,\theta) = \frac{2|D|}{n(n-1)} \le \frac{4|D|}{n^2} \le \frac{4\varepsilon_n^2}{t^2} + 8tf_{\max} + 4C_1\sqrt{\log(1/\delta)}n^{-1/2}$$

which is the claimed bound. We choose t to make the first to term equal, that is, set $t = (\frac{\varepsilon_n^2}{2f_{\max}})^{1/3}$, and noting that $16/2^{1/3} \le 13$ finishes the proof. The order of the bound is $(\varepsilon_n f_{\max})^{2/3}$.

C.3. Proofs of Lemma B.8 and Lemma B.6

Proof. Fix i,j with $i \sim j$. Let us expand $\gamma(\theta_j^*)$ and $\gamma'(\theta_j^*)$ around θ_i^*. To simplify, write $\delta = \theta_j^* - \theta_i^*$.

$$\begin{aligned}\gamma(\theta_j^*) &= \gamma(\theta_i^*) + \gamma'(\theta_i^*)\delta + \frac{1}{2}g\delta^2 \\ \gamma'(\theta_j^*) &= \gamma'(\theta_i^*) + \widetilde{g}\delta \end{aligned} \tag{21}$$

where g and $\widetilde{g}$ are $\gamma''(\cdot)$ evaluated at points between θ_i^* and θ_j^*. We note that $\|g\|$ and $\|\widetilde{g}\|$ are both $\le \kappa$. Adding $v_i^* = \gamma'(\theta_i^*)$ to the second equation and dividing by 2 we have

$$\begin{aligned} x_j^* - x_i^* &= v_i^*\delta + \frac{1}{2}g\delta^2 \\ \frac{1}{2}(v_j^* + v_i^*) &= v_i^* + \frac{1}{2}\widetilde{g}\delta. \end{aligned}$$

Multiplying we obtain,

$$W_{ij}^* = \delta + \frac{1}{2}\langle v_i^*, \widetilde{g}\rangle\delta^2 + \frac{1}{2}\langle v_i^*, g\rangle\delta^2 + \frac{1}{4}\langle g, \widetilde{g}\rangle\delta^3.$$

Note that $|\langle v_i^*, \widetilde{g}\rangle| \le \kappa$ and $|\langle v_i^*, g\rangle| \le \kappa$ and $|\langle g, \widetilde{g}\rangle| \le \kappa^2$. Since $|\delta| \le h^*$, we have

$$|W_{ij}^* - (\theta_j^* - \theta_i^*)| \le \kappa(h^*)^2 + \frac{1}{4}\kappa^2(h^*)^3$$

which proves Lemma B.8.

For Lemma B.6, from (21), we have $\langle v_j^*, v_i^*\rangle = 1 + \langle \widetilde{g}, v_i^*\rangle\delta$. But then, using $|\langle \widetilde{g}, v_i^*\rangle| \le \kappa$ and $|\delta| \le h^* \le 2\rho + r$ (see the argument after the statement of Lemma B.8), we obtain

$$\langle v_j^*, v_i^*\rangle \ge 1 - \kappa(2\rho + r) \ge 1/2$$

using the assumptions $\kappa(2\rho + r) \ll 1$. The proof is complete. □

C.4. Proof of Lemma B.9

For the first claim, defining $z = y - \theta + \bar{\theta}\mathbf{1}$, we see that y solves (11) iff z solves the same problem with $\widehat{W}$ replaced with Δ. The result now follows from Lemma 3.2, since $z = L^+\eta$ with η is given in the statement.

For the second claim, let $\eta_i = \sum_j A_{ij}\Delta_{ij}$. Using $\widehat{y} = \theta - \bar{\theta} + L^+\eta$, we have

$$\|\widehat{y} - (\theta - \bar{\theta})\| \le \|L^+\|\|\eta\| \le \frac{\|\eta\|}{\lambda_2(L)}$$

Next, $|\eta_i| \le d_i\|\Delta\|_\infty$. It follows that $\frac{1}{n}\|\eta\|^2 \le d_{\text{avg},2}^2\|\Delta\|_\infty^2$. Thus,

$$\frac{1}{n}\|\widehat{y} - (\theta - \bar{\theta})\|^2 \le \left(\frac{d_{\text{avg},2}}{\lambda_2(L)}\right)^2\|\Delta\|_\infty^2$$

Combined with Proposition B.5, the result follows.

D. Pseudo Code for STAGE

Algorithm 1 STAGE: Spectral Tangent Alignment and Geometric Embedding for Linear Seriation

Input: Data points $\mathscr{X}_n = \{x_1, \dots, x_n\} \subset \mathbb{R}^d$; neighborhood radius $r > 0$; minimum neighbors $d_{\min} \ge 2$.
Output: Permutation ϕ of $\{1, \dots, n\}$ representing the linear order.

```
/* Stage 1:  Local Tangent Estimation                                  */
```

foreach $x_i \in \mathscr{X}_n$ **do**
  Define neighborhood $\mathcal{N}(x_i, r) = \{x_j \in \mathscr{X}_n \mid \|x_j - x_i\|_2 \le r\}$.
  Let $d_i = |\mathcal{N}(x_i, r)|$.
  if $d_i \ge d_{\min}$ **then**
    Compute local mean: $\widehat{\mu}_i = \frac{1}{d_i}\sum_{x_j \in \mathcal{N}(x_i,r)} x_j$.
    Compute local covariance: $\widehat{\Sigma}_i = \frac{1}{d_i}\sum_{x_j \in \mathcal{N}(x_i,r)}(x_j - \widehat{\mu}_i)(x_j - \widehat{\mu}_i)^\intercal$.
    $\widehat{v}_i \leftarrow$ normalized principal eigenvector of $\widehat{\Sigma}_i$.
  else
    $\widehat{v}_i \leftarrow \mathbf{0}$ (or handle as missing/unreliable). *// Tangent undefined*

```
/* Stage 2:  Orientation Alignment                                     */
```

Construct affinity matrix $\widehat{Q} \in \mathbb{R}^{n\times n}$ where:

$$\widehat{Q}_{ij} = \begin{cases} \langle \widehat{v}_i, \widehat{v}_j \rangle & \text{if } i \sim j \text{ and } \widehat{v}_i, \widehat{v}_j \ne \mathbf{0} \\ 0 & \text{otherwise} \end{cases} \qquad // i \sim j \iff \|x_i - x_j\|_2 \le r$$

Compute $\widehat{u} \leftarrow$ principal eigenvector of $\widehat{Q}$.
foreach $i \in \{1, \dots, n\}$ **do**
  $\widehat{s}_i \leftarrow \operatorname{sign}(\widehat{u}_i)$ (or 1 if $\widehat{u}_i = 0$ or $\widehat{v}_i = \mathbf{0}$).
Let $\widetilde{v}_i = \widehat{s}_i\widehat{v}_i$ be the oriented tangents.

```
/* Stage 3:  Global Embedding via Least Squares                        */
```

foreach *pair* (i, j) *such that* $i \sim j$ **do**
  $\widehat{W}_{ij} \leftarrow \frac{1}{2}(\widetilde{v}_i + \widetilde{v}_j)^\intercal(x_j - x_i)$.
Solve for the embedding $\widehat{y} \in \mathbb{R}^n$:

$$\widehat{y} = \operatorname*{argmin}_{y \in \mathbb{R}^n:\ \langle \mathbf{1}, y\rangle = 0} \sum_{i,j:\ i\sim j} \left(y_j - y_i - \widehat{W}_{ij}\right)^2.$$

```
/* Stage 4:  Final Ordering                                            */
```

$\phi \leftarrow \operatorname{argsort}_{j\in\{1,\dots,n\}}\widehat{y}_j$.
return ϕ

E. Additional Experiments

E.1. 2-D Experiments

Our 2-D dataset consists of points sampled from a rotated sine curve embedded in $\mathbb{R}^2$. We introduce additive Gaussian noise with standard deviation σ varying from 0.1 to 0.5 (in increments of 0.1). Experiments were run for sample sizes

$n \in \{500, 1000, 3000\}$. For each configuration (noise level and sample size), we conducted 10 independent trials, reporting the average Kendall's τ and its standard deviation in Table 3.

The results in Table 3 demonstrate that STAGE consistently achieves the highest absolute Kendall's τ values across all tested noise levels and sample sizes, indicating superior order recovery compared to the baseline methods. For clarity, Kendall's τ values in Table 3 are multiplied by 100.

Table 3. Comparison of Methods on 2-D sine curve, Kendall's τ and std. err. values multiplied by 100

Sample	Method	$\sigma = 0.1$	$\sigma = 0.2$	$\sigma = 0.3$	$\sigma = 0.4$	$\sigma = 0.5$
500	Fiedler Vector	61.39 (5.06)	73.0 (5.51)	74.1 (5.89)	74.0 (6.2)	76.55 (3.83)
	1D UMAP	70.59 (13.25)	69.35 (13.08)	71.21 (16.11)	66.32 (15.94)	73.52 (7.34)
	1D t-SNE	97.82 (0.11)	95.38 (0.21)	92.76 (0.21)	90.17 (0.41)	87.2 (0.75)
	Recanati	97.65 (0.1)	95.35 (0.18)	92.46 (0.43)	87.92 (1.61)	81.54 (5.34)
	STAGE	**97.85 (0.04)**	**95.41 (0.20)**	**93.03 (0.20)**	**90.26 (0.38)**	**87.34 (0.39)**
1000	Fiedler Vector	65.63 (2.85)	74.98 (3.94)	78.31 (3.79)	76.8 (1.81)	68.4 (22.68)
	1D UMAP	73.68 (12.83)	77.33 (11.5)	73.19 (14.24)	67.67 (14.99)	74.71 (12.53)
	1D t-SNE	97.69 (0.09)	95.2 (0.15)	92.56 (0.11)	90.0 (0.37)	87.08 (0.49)
	Recanati	97.64 (0.08)	95.23 (0.11)	92.47 (0.35)	84.53 (5.0)	78.57 (3.2)
	STAGE	**97.82 (0.06)**	**95.47 (0.09)**	**93.03 (0.14)**	**90.34 (0.32)**	**87.51 (0.28)**
3000	Fiedler Vector	66.49 (3.67)	77.49 (2.54)	78.06 (3.03)	78.8 (2.31)	78.16 (1.44)
	1D UMAP	80.88 (7.65)	79.99 (8.43)	86.69 (4.82)	81.6 (5.76)	75.27 (5.54)
	1D t-SNE	97.3 (0.38)	95.08 (0.12)	92.56 (0.11)	89.88 (0.22)	87.21 (0.22)
	Recanati	97.6 (0.02)	95.22 (0.09)	92.32 (0.69)	85.11 (2.96)	75.04 (3.04)
	STAGE	**97.76 (0.04)**	**95.44 (0.07)**	**92.92 (0.11)**	**90.30 (0.18)**	**87.54 (0.22)**

E.2. Higher Dimension Experiments

Here, we provide additional experiments to demonstrate the performance of the STAGE algorithm. In Figure 5 we display an example random Fourier curve in 3 dimensions. Table 4 shows the full results of our high-dimensional experiments, across a range of sample sizes. Figure 6 shows the pairplots as shown in Figure 2 in the main text, along with the true curves to demonstrate the effectiveness of the denoising. In Figures 7 and 8 we demonstrate the increase in performance and runtime of STAGE and t-SNE as the neighborhood size and perplexity increase, respectively. These plots clearly show both the superior accuracy of STAGE and the lower computational cost relative to t-SNE. We perform similar comparisons to additional embedding methods ISOMAP and LLE in Figures 9, 10 and 11. We compare the methods with $n = 500, 1000, 2000$ and $d = 50, 100, 200$, with 10 replicates per setting and $\sigma = 2$. These experiments show that we do somewhat better at recovery than ISOMAP, particularly at higher noise levels and dimensionality and greatly outperform LLE. LLE is faster (but lacks accuracy), whereas STAGE is slightly slower than ISOMAP at lower sample sizes, but faster at $n = 2000$. LLE also seems much more sensitive to the choice of neighborhood size. We further explore the differences in compute time with respect to all the methods in Table 5, where we see that our method is comparable to UMAP, roughly 2× faster than arguably the strongest competitor, t-SNE ($\sim$2.3× at $n = 1000$, $\sim$2.2× at $n = 2000$), and $\sim$3× faster than Recanati and Fiedler at $n = 1000$, rising to $\sim$4–5× at $n = 2000$. Finally, in order to better understand the effect of curvature and noise on our model, we apply our method to unit-length arcs drawn from circles whose curvatures range from 0 to 5, and with σ varying between 0 and 5. We ran these simulations for $n = 1000$ and $n = 2000$, where for each combination of σ and κ we embedded the circle in $d = 10, 50, 500$ and averaged the results of 20 replicates. In Figure 12, we provide a heatmap of the average τ over κ and σ where we see that as κ increases or as d increases, so does the sensitivity to noise.

E.3. Related Tangent-Based Manifold Methods

Several related manifold-learning methods also begin by estimating local tangent geometry, but their global objectives differ from STAGE's seriation objective. LTSA (Zhang & Zha, 2004) estimates local tangent bases Q_i and local coordinates, then seeks global coordinates T together with unconstrained local linear maps L_i. Eliminating the L_i gives a homogeneous spectral problem for an alignment matrix, with an orthogonality constraint used to exclude the trivial solution. In one dimension, the free scalar L_i can absorb the sign ambiguity of the local tangent estimate. STAGE instead resolves

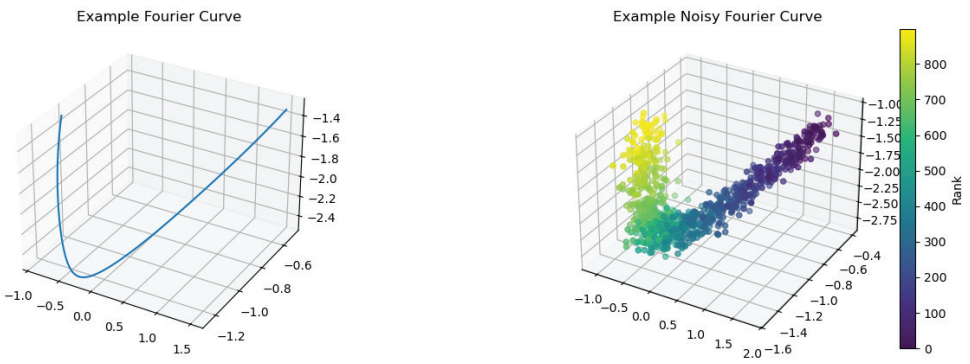


Figure 5. (Left) First 3 dimensions of true Fourier curve, (Right) First 3 dimensions of perturbed Fourier curve, colored by order

this ambiguity explicitly by a graph-wide $\mathbb{Z}_2$ synchronization step, then constructs signed edgewise regression targets $\widehat{W}_{ij} = \frac{1}{2}(\widehat{s}_i\widehat{v}_i + \widehat{s}_j\widehat{v}_j)^\top(x_j - x_i)$. These targets directly estimate local increments of the latent parameter and lead to an inhomogeneous least-squares problem with a unique centered solution. This is also why the Laplacian in STAGE appears as a normal-equation matrix, not as the spectral embedding operator itself. The distinction is consistent with the empirical LTSA comparison in Table 7.

VDM (Singer & Wu, 2012) and LPD (Wolf & Averbuch, 2013), together with the finite-sample Patch-to-Tensor Embedding construction (Salhov et al., 2012), are closer to STAGE in that they aggregate local tangent-space relationships. VDM forms orthogonal inter-tangent alignments and uses them to build a connection matrix whose spectral embedding yields vector-diffusion distances. LPD uses raw projection blocks $O_i^\top O_j$ in a super-kernel, retaining the angle between tangent spaces. For intrinsic dimension one, an LPD block is algebraically close to a weighted version of STAGE's orientation matrix $\widehat{Q}_{ij} = \langle \widehat{v}_i, \widehat{v}_j \rangle \mathbf{1}\{i \sim j\}$. The usage is different: STAGE uses $\widehat{Q}$ to synchronize tangent signs before constructing scalar increments, while LPD feeds the projection blocks directly into a spectral diffusion construction. Neither VDM nor LPD/PTE returns a one-dimensional ordering by itself, and neither uses ambient displacements projected onto oriented tangents to form regression targets. Extracting a seriation from their distances would require an additional ordering step, again passing through pairwise distance information.

It is worth noting that the initial local PCA step shared by STAGE and many other methods is not new. It is a standard tool in manifold learning, and local SVD/PCA estimates for curves have also been studied directly (Álvarez Vizoso et al., 2019). The novelty of STAGE is the subsequent steps of explicit sign synchronization and displacement-based regression leading to an order-preserving emebedding.

E.4. Additional Comparisons and Ablations

To further address reviewer questions, we include here three additional comparisons: a comparison with LTSA, an ablation in which the tangent-alignment step is removed, and a numerical comparison between the direct LinReg solve and its Laplacian form. For compactness, we report representative high-dimensional results for $n = 500$; the same qualitative conclusions persisted for larger sample sizes in our additional runs. We additionally compare with SABRE (Issartel et al., 2024). Like Recanati's method, SABRE starts from a similarity matrix rather than the raw point cloud; for this comparison we construct the input similarity matrix from the observed coordinates using a Gaussian kernel on pairwise Euclidean distances. Table 6 reports the resulting order recovery and runtime.

E.5. scRNA-Seq Data

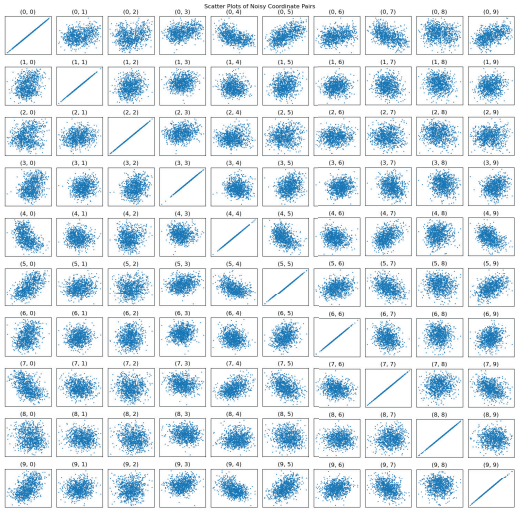
Here we include some further analysis of the scRNA-Seq datasets discussed in the main paper. In particular, we show the distribution of the residuals and a degree analysis of the resulting neighborhood graphs. In Figures 13, 14 and 15 we see that for the most part, the residuals look normal. There is some deviation in the Packer seam cell data, which has some heavier tails, but nonetheless our performance for this dataset is quite strong. Finally, we tested for heavy tails by fitting a power-law tail and comparing it to thin-tailed alternatives using the Clauset–Shalizi–Newman procedure (Clauset

et al., 2009) (Python powerlaw, discrete degrees). The method (i) selects x_{min} by minimizing the KS distance, (ii) fits the power-law exponent on $d \geq x_{min}$, and (iii) uses log-likelihood ratio (LLR) tests to compare power-law vs log-normal and exponential tails. The results can be seen in Table 10. Negative LLR means the alternative fits better; the reported p is the p-value for that comparison. We also summarize overall concentration via CV (SD/mean) and the 99th-percentile/median ratio.

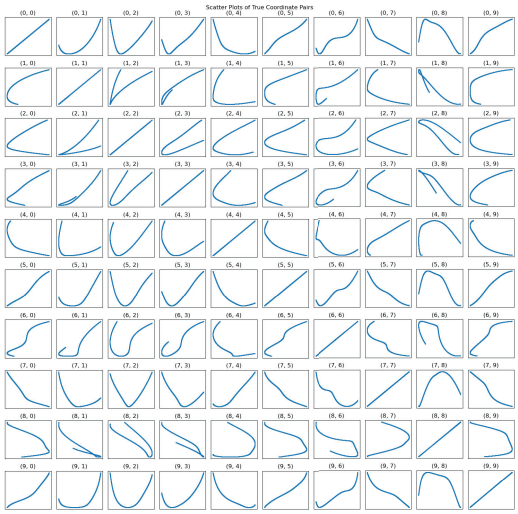
We also include an analysis of results based on different factors of the smallest connectivity radius for each dataset (denoted r_0) in Figure 16. These demonstrate that for some factor of the minimal connectivity radius, we get good results with STAGE, and past that point the performance drops off, but not significantly. The behavior is not as well demonstrated in the last dataset - a direction for further investigation might be to understand this behavior.

F. Large Language Model (LLM) Usage

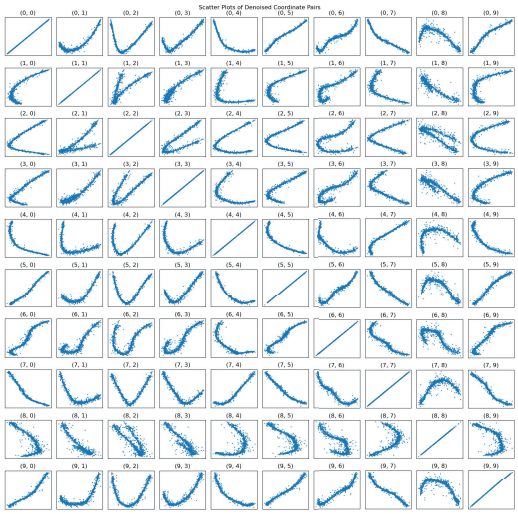
We used LLMs (GPT-5 and Gemini) to help with writing code and polishing the writing of the paper.



(a) Noisy data



(b) True curve



(c) Denoised (STAGE)

Figure 6. Pairplots of the first 10 dimensions: (a) original noisy observations, (b) underlying curve, and (c) denoised points produced by STAGE.

Table 4. Comparison of Methods on high-dimensional Fourier curves, Kendall's τ and std. err. values multiplied by 100

Dim.	Sample	Method	$\sigma = 0.5$	$\sigma = 1$	$\sigma = 1.5$	$\sigma = 2$
50	500	Fiedler Vector	81.7 (8.62)	77.55 (3.51)	60.03 (7.39)	28.82 (17.16)
		1D UMAP	88.27 (0.97)	75.08 (2.62)	60.42 (4.55)	38.93 (14.45)
		1D t-SNE	90.62 (0.87)	75.53 (15.43)	68.52 (3.66)	57.77 (5.52)
		Recanati	89.86 (1.14)	79.31 (1.67)	70.11 (1.85)	62.95 (2.21)
		STAGE	**91.11 (0.9)**	**82.06 (1.59)**	**72.28 (2.19)**	**63.79 (2.73)**
	1000	Fiedler Vector	84.58 (6.86)	77.45 (3.21)	44.81 (18.28)	26.89 (8.16)
		1D UMAP	88.43 (1.0)	76.02 (2.0)	63.1 (3.81)	47.59 (6.24)
		1D t-SNE	90.42 (0.79)	80.6 (1.58)	71.36 (1.85)	60.71 (4.97)
		Recanati	89.87 (1.2)	79.61 (1.38)	71.96 (1.76)	63.53 (2.81)
		STAGE	**90.92 (0.78)**	**81.69 (1.44)**	**73.11 (1.91)**	**64.06 (3.4)**
	2000	Fiedler Vector	85.47 (4.53)	76.91 (3.53)	49.49 (11.23)	33.22 (14.77)
		1D UMAP	88.6 (0.67)	77.08 (1.89)	64.5 (3.41)	51.17 (3.5)
		1D t-SNE	90.51 (0.63)	81.03 (1.43)	71.53 (2.75)	61.51 (3.16)
		Recanati	90.05 (1.03)	80.1 (1.35)	71.6 (1.86)	**64.39 (2.35)**
		STAGE	**90.91 (0.66)**	**81.76 (1.47)**	**72.26 (2.36)**	64.04 (2.58)
100	500	Fiedler Vector	75.88 (12.19)	82.34 (1.84)	69.3 (3.69)	48.96 (8.28)
		1D UMAP	90.66 (0.66)	80.5 (1.31)	66.4 (2.19)	55.4 (4.61)
		1D t-SNE	92.89 (0.43)	69.86 (23.17)	75.24 (4.65)	65.74 (6.4)
		Recanati	92.32 (0.59)	83.41 (0.65)	76.55 (0.92)	70.23 (1.81)
		STAGE	**93.42 (0.39)**	**86.46 (0.72)**	**79.23 (1.18)**	**72.74 (2.27)**
	1000	Fiedler Vector	82.12 (9.06)	82.06 (2.84)	66.65 (6.98)	41.75 (18.69)
		1D UMAP	90.93 (0.75)	80.82 (1.5)	70.61 (1.56)	55.6 (6.88)
		1 -D t-SNE	92.91 (0.49)	81.58 (10.42)	78.03 (1.76)	67.84 (4.79)
		Recanati	92.55 (0.66)	84.11 (0.75)	77.55 (0.94)	71.61 (1.08)
		STAGE	**93.41 (0.45)**	**86.47 (0.72)**	**79.59 (1.23)**	**72.38 (1.54)**
	2000	Fiedler Vector	83.1 (12.9)	83.34 (1.55)	63.0 (13.29)	44.9 (11.3)
		1D UMAP	90.89 (0.45)	81.72 (1.18)	71.14 (2.07)	58.85 (4.81)
		1D t-SNE	92.93 (0.33)	85.66 (0.86)	78.28 (1.49)	70.05 (3.35)
		Recanati	92.77 (0.47)	84.51 (0.76)	77.9 (1.03)	72.07 (1.39)
		STAGE	**93.37 (0.29)**	**86.46 (0.85)**	**79.47 (1.39)**	**72.27 (1.89)**
200	500	Fiedler Vector	40.39 (15.15)	86.18 (4.95)	79.75 (3.77)	65.57 (9.33)
		1D UMAP	93.17 (0.31)	85.89 (0.7)	77.18 (1.39)	66.43 (2.57)
		1D t-SNE	95.18 (0.19)	89.73 (0.58)	77.01 (18.78)	76.19 (3.27)
		Recanati	94.89 (0.21)	87.17 (0.43)	82.1 (0.92)	76.87 (0.87)
		STAGE	**95.63 (0.22)**	**90.99 (0.27)**	**86.22 (0.66)**	**80.98 (0.84)**
	1000	Fiedler Vector	62.64 (30.52)	84.71 (9.72)	81.12 (2.29)	64.05 (6.77)
		1D UMAP	85.22 (23.88)	85.86 (0.88)	78.01 (0.75)	68.86 (1.46)
		1D t-SNE	95.13 (0.17)	89.96 (0.32)	73.14 (20.3)	76.26 (4.62)
		Recanati	95.13 (0.16)	87.62 (0.42)	82.92 (0.6)	78.12 (0.55)
		STAGE	**95.58 (0.14)**	**90.95 (0.36)**	**86.03 (0.55)**	**81.31 (0.52)**
	2000	Fiedler Vector	68.8 (22.04)	88.7 (0.37)	81.2 (1.1)	66.52 (4.1)
		1D UMAP	93.29 (0.33)	86.35 (0.46)	78.89 (0.62)	70.17 (1.03)
		1D t-SNE	95.01 (0.08)	89.94 (0.26)	84.98 (0.28)	77.47 (4.18)
		Recanati	95.17 (0.12)	87.84 (0.21)	83.41 (0.34)	78.88 (0.49)
		STAGE	**95.47 (0.09)**	**90.77 (0.24)**	**86.12 (0.28)**	**81.18 (0.56)**

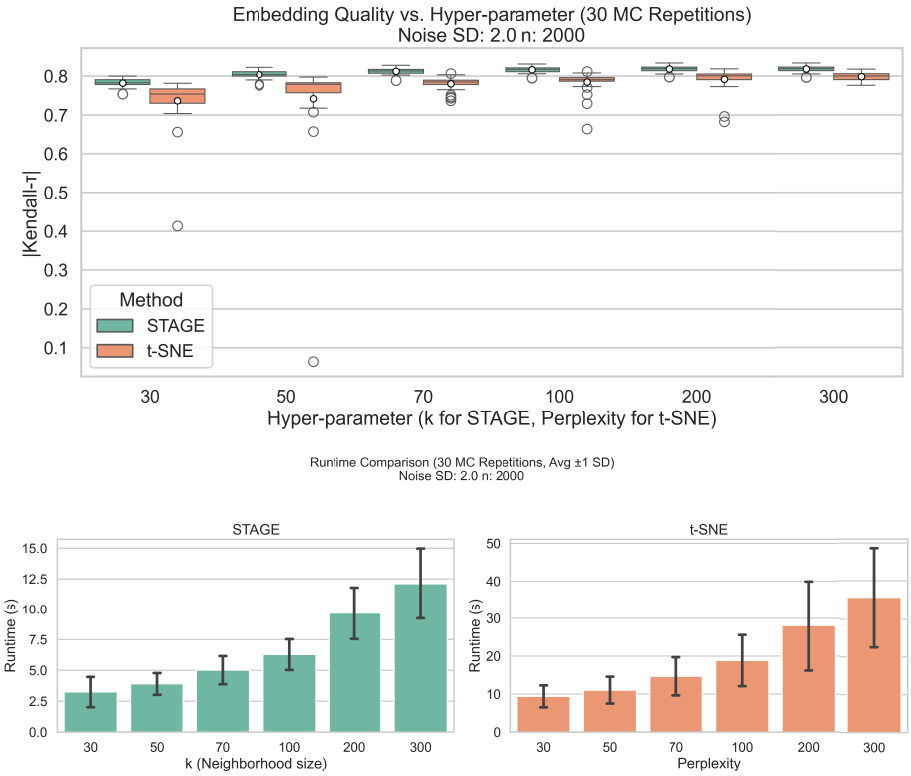


Figure 7. (Top) Order recovery of STAGE vs. t-SNE as neighborhood size/perplexity increases, (Bottom) Runtime of STAGE vs. t-SNE as neighborhood size/perplexity increases (y-axis scales differ)

Table 5. Comparison of runtimes (in seconds) on high-dimensional Fourier curves

Dim.	Sample	Fiedler Vector	1D UMAP	1D t-SNE	Recanati	STAGE
50	500	0.4651	0.5631	0.7526	0.5212	0.2468
	1000	1.9515	0.5643	1.8137	1.775	0.4894
	2000	8.7202	1.5007	2.5898	6.8389	1.1483
100	500	0.497	0.2493	0.7831	0.5239	0.2769
	1000	2.0782	0.5483	1.5271	1.8017	0.5879
	2000	9.5623	1.6735	3.5941	7.2733	1.4597
200	500	0.4798	0.2378	1.5625	0.5283	0.3081
	1000	2.1054	0.6303	1.5625	1.8157	0.6904
	2000	8.7893	1.5472	3.605	6.8337	1.6502

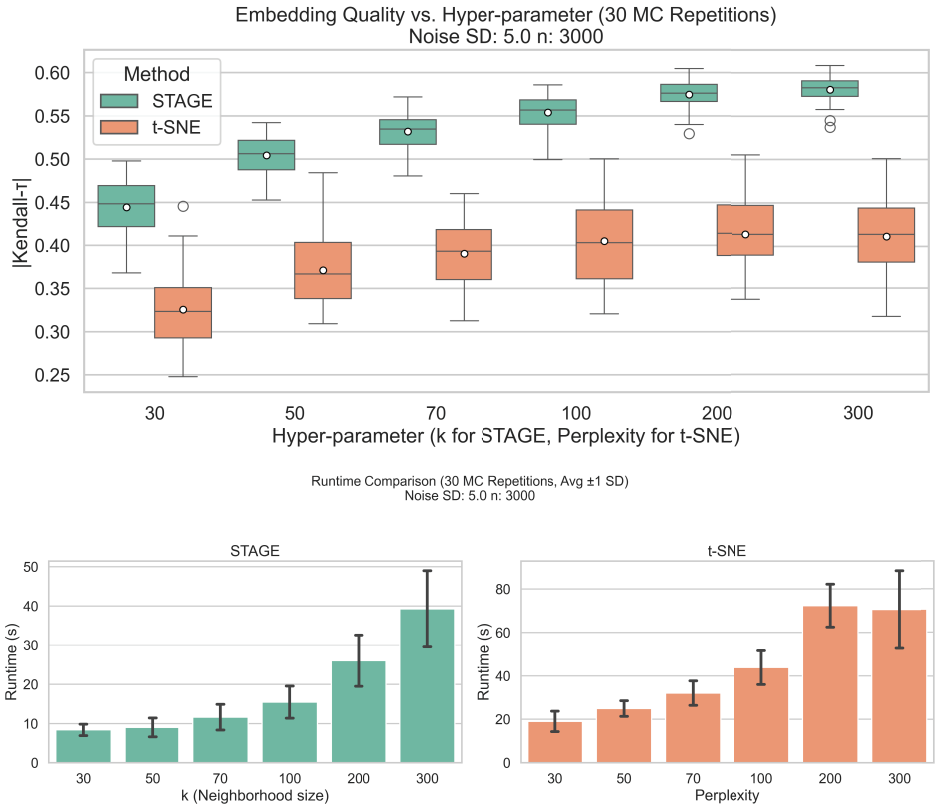


Figure 8. (Top) Order recovery of STAGE vs. t-SNE as neighborhood size/perplexity increases, (Bottom) Runtime of STAGE vs. t-SNE as neighborhood size/perplexity increases (y-axis scales differ)

Table 6. STAGE vs. SABRE on high-dimensional Fourier curves ($N = 500$). Mean $|\tau| \times 100$ with standard deviation in parentheses; runtime entries are mean seconds.

Dim.	Method	$\sigma = 0.5$	$\sigma = 1.0$	$\sigma = 1.5$	$\sigma = 2.0$
Mean $\lvert\tau\rvert \times 100$					
50	STAGE	90.95 (0.79)	82.01 (1.27)	72.19 (2.69)	63.26 (4.30)
	SABRE	58.92 (31.57)	13.18 (20.17)	14.83 (10.19)	5.04 (5.08)
100	STAGE	93.45 (0.40)	86.97 (0.72)	79.49 (1.26)	72.06 (1.75)
	SABRE	57.56 (35.81)	6.00 (7.91)	8.29 (11.02)	7.42 (7.14)
200	STAGE	95.72 (0.19)	90.95 (0.34)	86.37 (0.55)	81.30 (0.77)
	SABRE	78.18 (23.10)	60.03 (36.80)	9.34 (11.34)	11.12 (12.24)
Mean runtime (seconds)					
50	STAGE	0.231	0.249	0.248	0.264
	SABRE	1.200	1.087	0.925	0.849
100	STAGE	0.270	0.274	0.281	0.273
	SABRE	1.277	1.152	0.996	0.867
200	STAGE	0.277	0.286	0.294	0.297
	SABRE	1.179	1.113	1.022	0.913

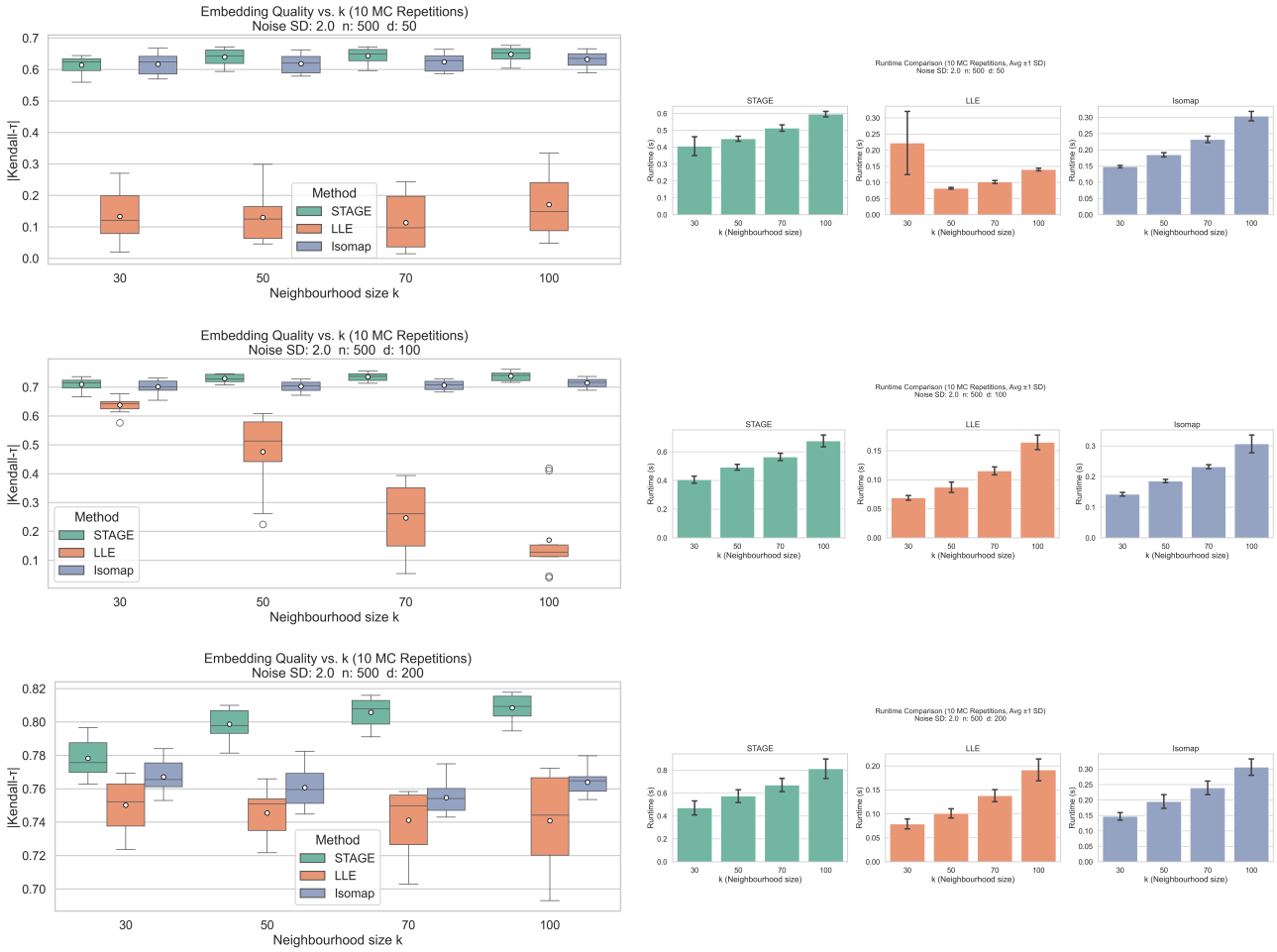


Figure 9. Kendall's τ and runtime comparison of STAGE vs. ISOMAP vs. LLE with n = 500

Table 7. STAGE vs. LTSA on high-dimensional Fourier curves ($n = 500$). Mean $|\tau| \times 100$ with standard deviation in parentheses.

Dim.	Method	$\sigma = 0.5$	$\sigma = 1.0$	$\sigma = 1.5$	$\sigma = 2.0$
50	STAGE	91.10 (0.73)	81.84 (2.05)	73.12 (2.19)	63.32 (3.58)
	LTSA	77.34 (23.85)	43.42 (27.42)	23.02 (20.05)	13.59 (13.47)
100	STAGE	93.38 (0.45)	86.53 (0.89)	79.55 (1.33)	72.43 (1.72)
	LTSA	86.64 (13.76)	39.39 (28.08)	13.34 (20.82)	13.22 (14.74)
200	STAGE	95.60 (0.16)	90.88 (0.26)	85.99 (0.61)	81.01 (0.60)
	LTSA	73.89 (32.43)	52.91 (37.02)	33.62 (29.83)	21.25 (21.60)

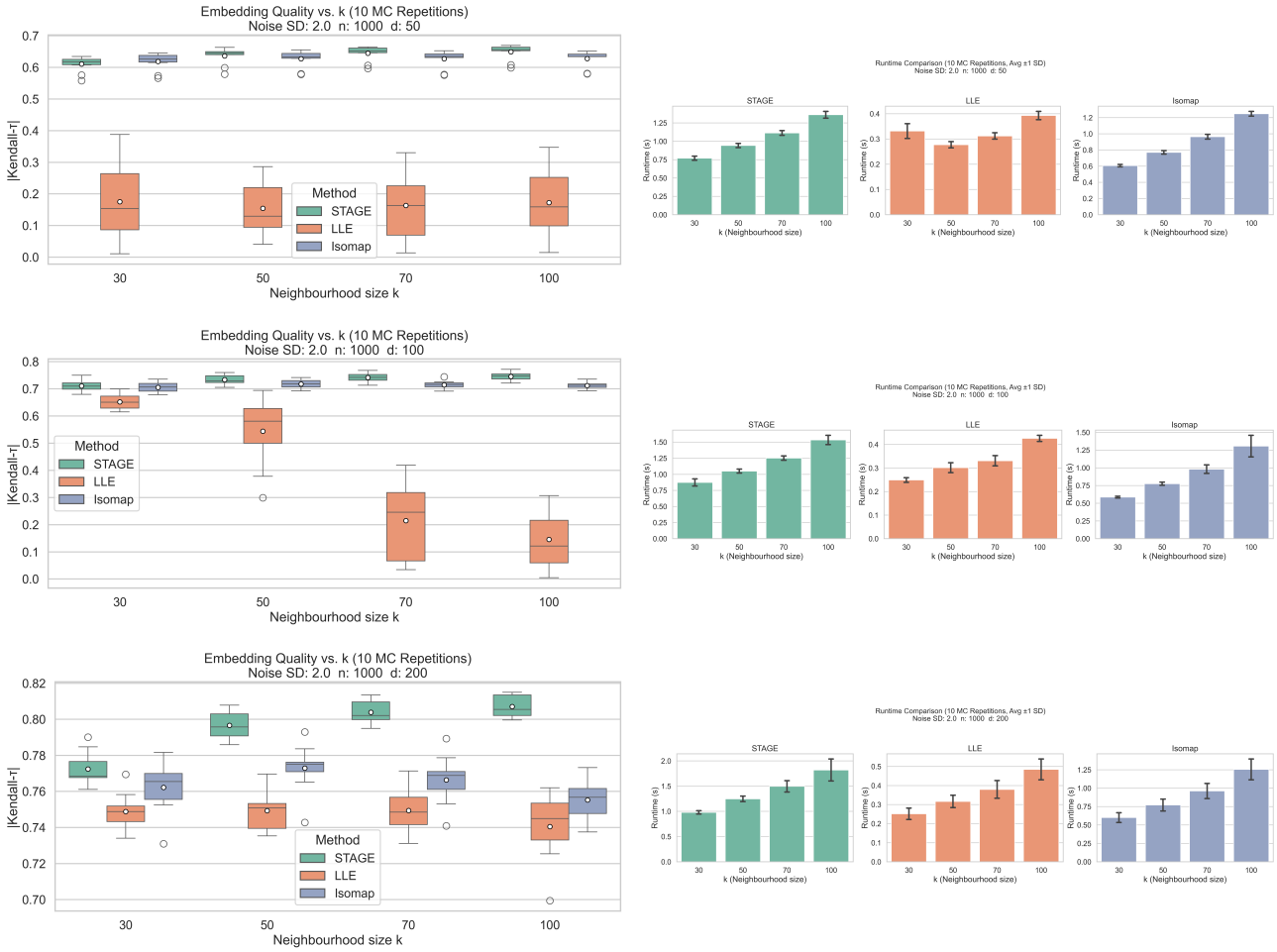


Figure 10. Kendall's τ and runtime comparison of STAGE vs. ISOMAP vs. LLE with n = 1000

Table 8. Ablation of the orientation-alignment step on high-dimensional Fourier curves ($n = 500$). Mean $|\tau| \times 100$ with standard deviation in parentheses.

Dim.	Method	$\sigma = 0.5$	$\sigma = 1.0$	$\sigma = 1.5$	$\sigma = 2.0$
50	STAGE	90.94 (0.75)	82.04 (1.63)	72.27 (2.64)	63.45 (3.09)
	STAGE-no-align	73.39 (20.66)	63.06 (23.63)	52.54 (19.78)	41.04 (18.71)
100	STAGE	93.44 (0.42)	86.65 (1.16)	79.36 (1.31)	72.12 (1.98)
	STAGE-no-align	32.49 (25.78)	41.07 (15.80)	37.69 (22.72)	41.54 (17.67)
200	STAGE	95.51 (0.18)	91.14 (0.36)	86.40 (0.64)	80.87 (0.64)
	STAGE-no-align	34.54 (15.46)	31.15 (10.45)	36.08 (20.59)	42.70 (15.35)

Table 9. Numerical agreement between the direct LinReg solve and the Laplacian form on high-dimensional Fourier curves ($n = 500$). Entries are mean relative RMSE between the two embeddings, with the maximum over 10 runs in parentheses. In every setting, the Kendall's τ between the two recovered orderings was exactly 1.0.

Dim.	$\sigma = 0.5$	$\sigma = 1.0$	$\sigma = 1.5$	$\sigma = 2.0$
50	9.721e-12 (1.485e-11)	2.565e-11 (3.373e-11)	5.998e-11 (1.265e-10)	8.587e-11 (1.528e-10)
100	9.279e-12 (1.663e-11)	2.268e-11 (3.816e-11)	5.229e-11 (9.823e-11)	7.783e-11 (1.287e-10)
200	6.211e-12 (9.067e-12)	1.582e-11 (2.236e-11)	2.818e-11 (3.979e-11)	4.948e-11 (6.732e-11)

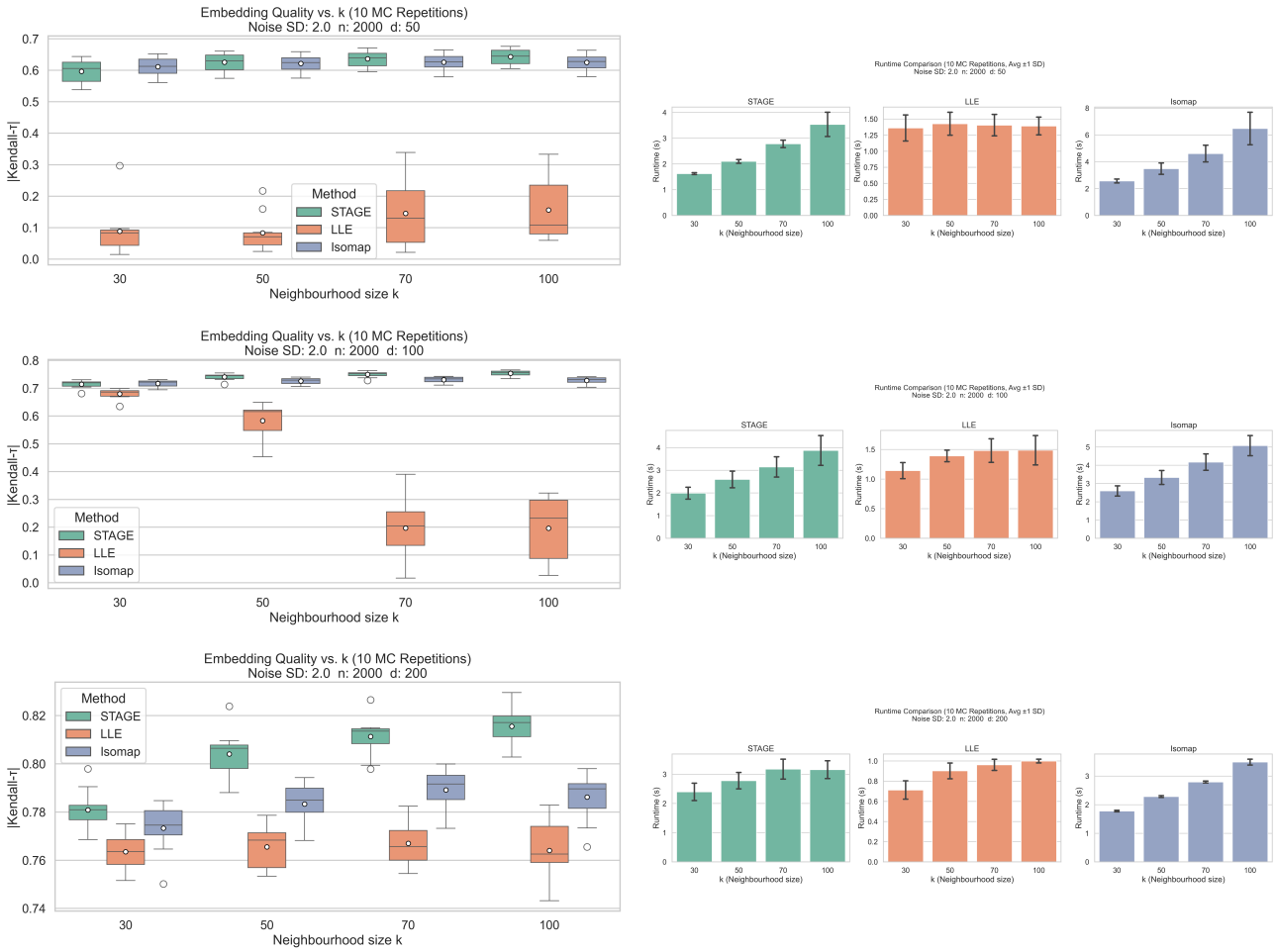


Figure 11. Kendall's τ and runtime comparison of STAGE vs. ISOMAP vs. LLE with n = 2000

Table 10. Tail tests & concentration (per dataset)

Dataset	LLR (vs LN)	p (vs LN)	LLR (vs exp)	p (vs exp)	CV	99th/median
Saelens	-17.67	6.77e-70	-29.01	5.45e-185	0.42	1.27
Packer Seam Cells	-15.51	2.86e-54	-16.78	3.48e-63	0.75	3.81
Packer Hypodermis	-37.34	3.33e-305	-79.19	0	0.50	1.38

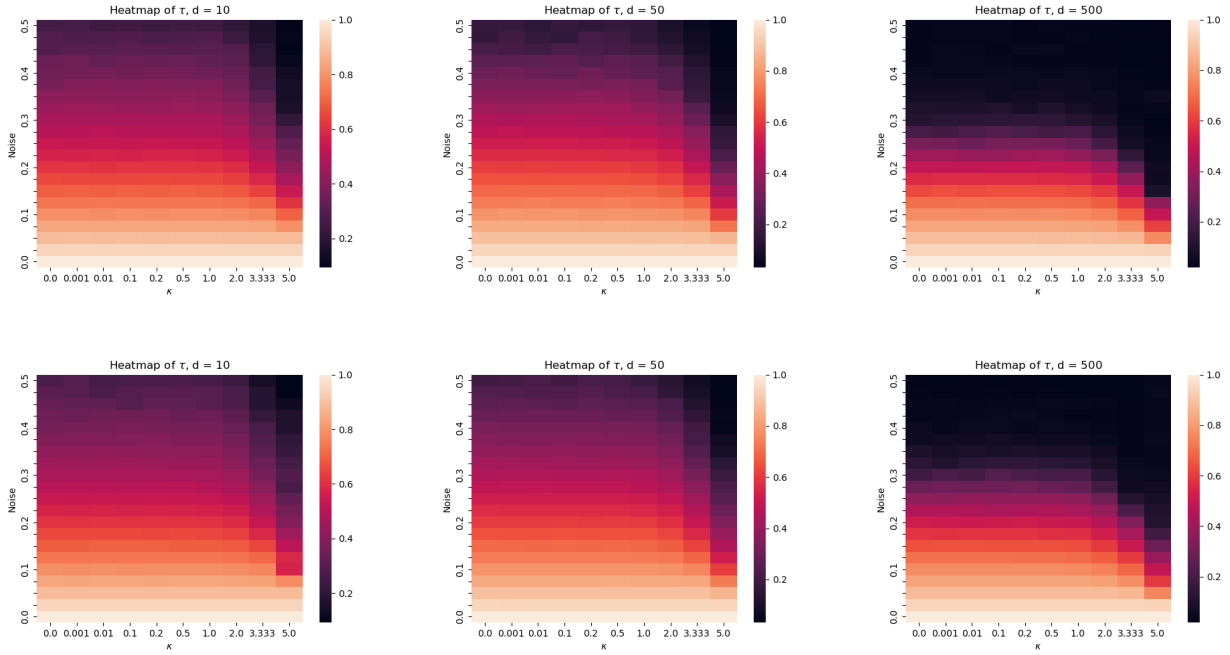
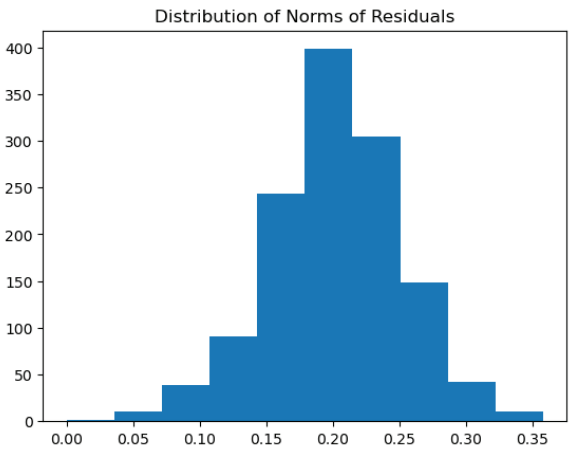
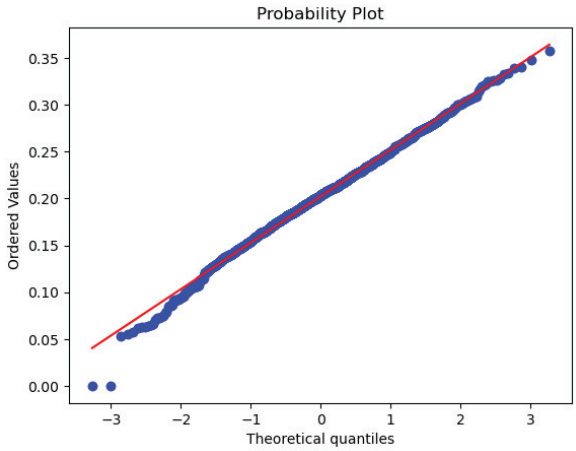


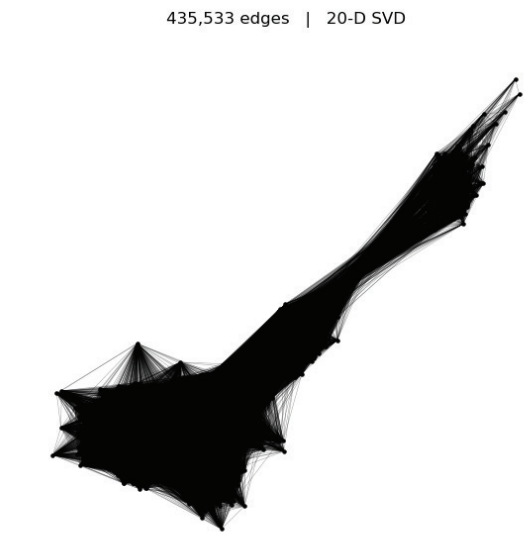
Figure 12. Heatmap of τ (averaged over 20 runs) for various values of κ and σ. Top row: $n = 1000$ with $d = 10, 50, 500$ from left to right. Bottom row: $n = 2000$ with $d = 10, 50, 500$ from left to right.



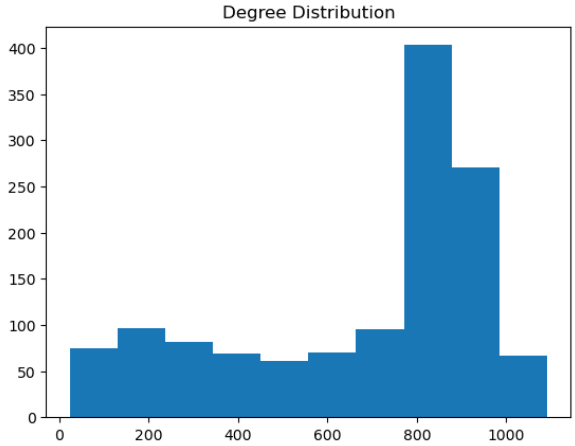
(a) Histogram of residuals



(b) Q-Q plot of residuals

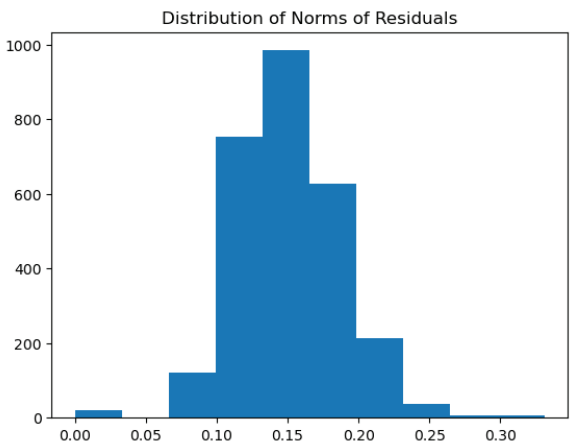


(c) Neighborhood graph

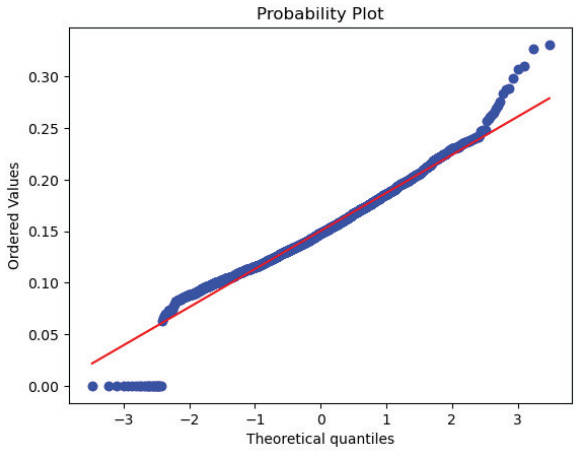


(d) Histogram of degrees

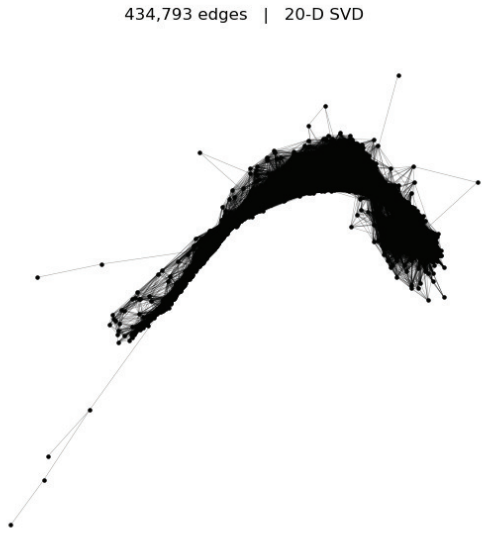
Figure 13. Residual and degree distributions of the STAGE algorithm applied to the Saelens dataset: (a) histogram of residuals, (b) Q-Q plot of residuals, (c) neighborhood graph, and (d) histogram of degrees



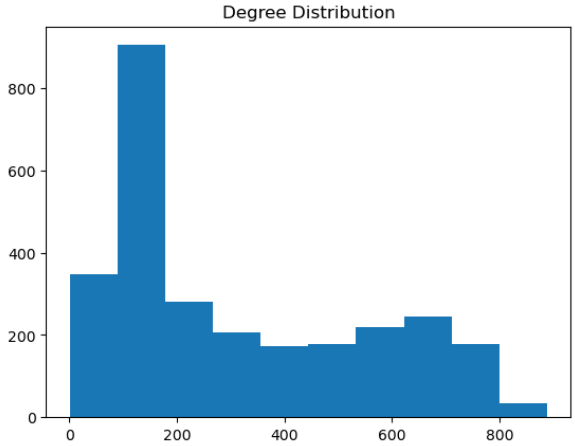
(a) Histogram of residuals



(b) Q-Q plot of residuals

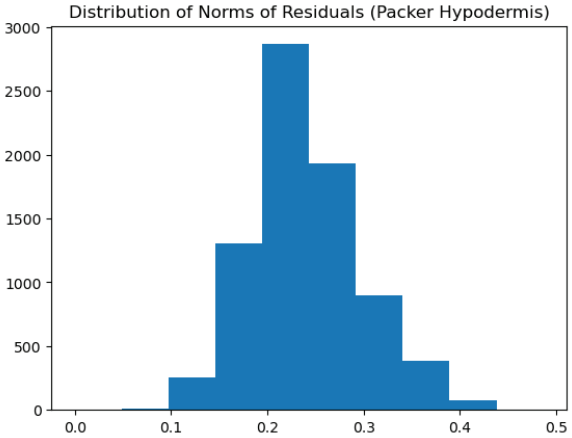


(c) Neighborhood graph

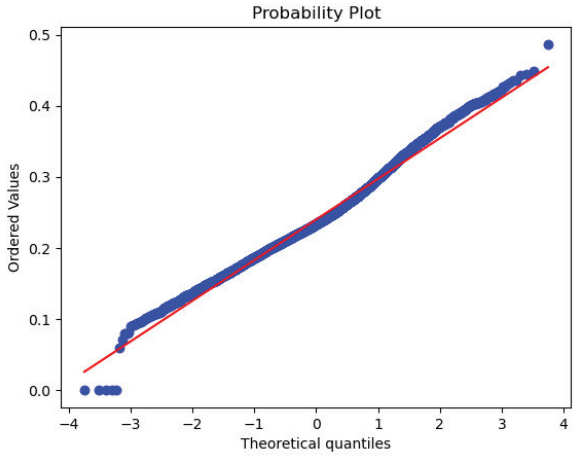


(d) Histogram of degrees

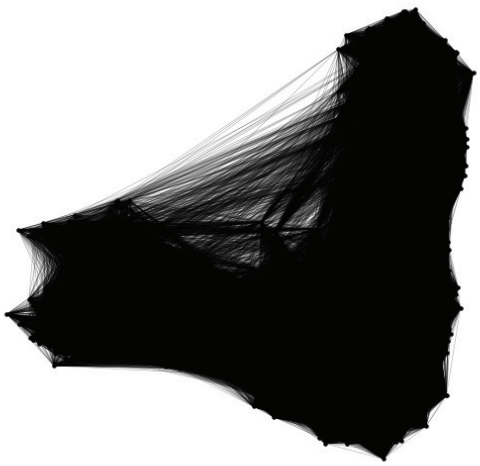
Figure 14. Residual and degree distributions of the STAGE algorithm applied to the Packer seam cell dataset: (a) histogram of residuals, (b) Q-Q plot of residuals, (c) neighborhood graph, and (d) histogram of degrees



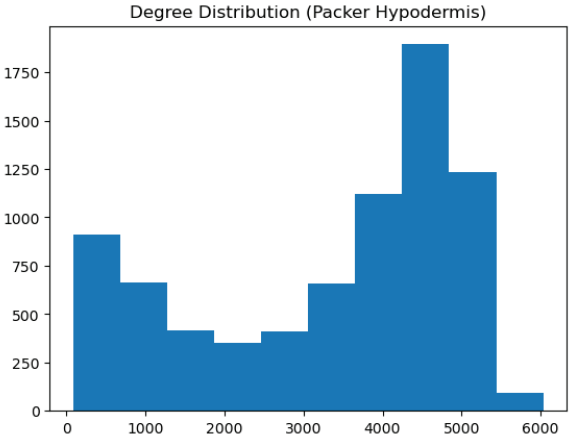
(a) Histogram of residuals



(b) Q-Q plot of residuals

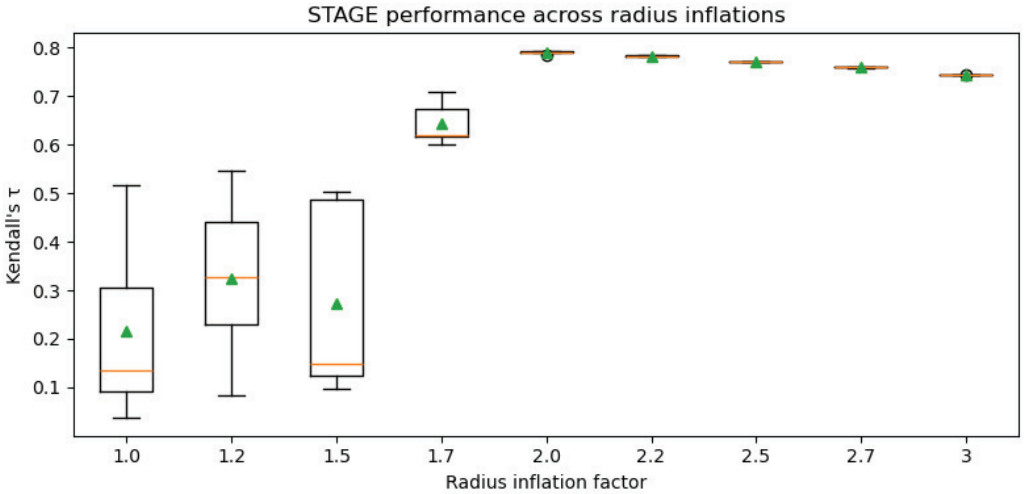


(c) Neighborhood graph

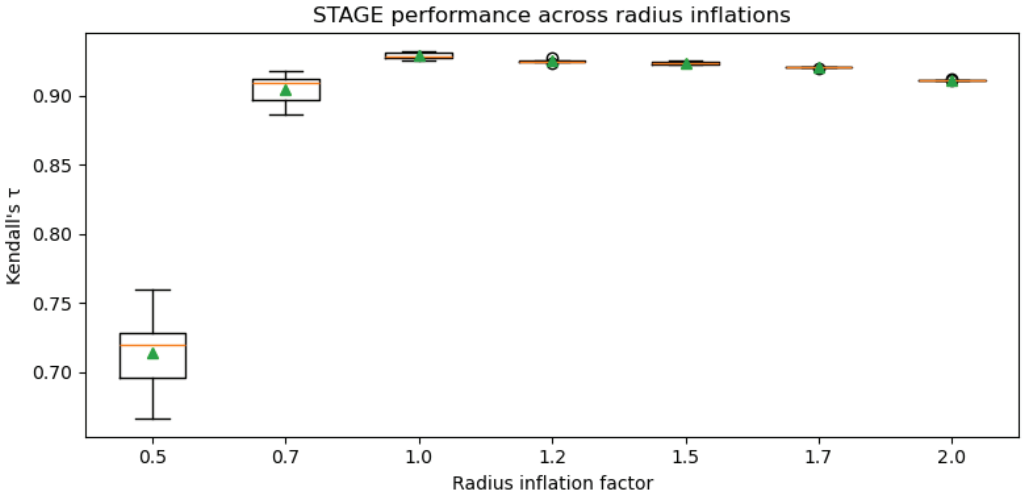


(d) Histogram of degrees

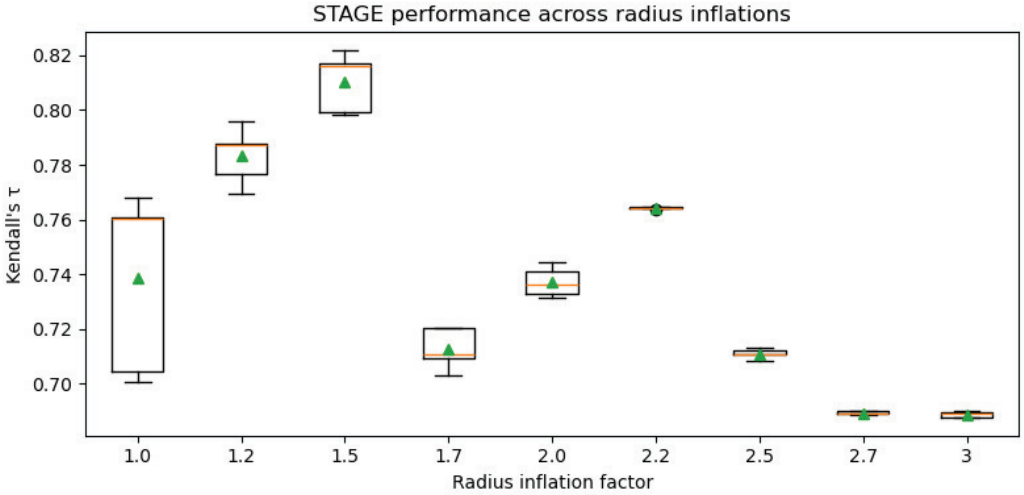
Figure 15. Residual and degree distributions of the STAGE algorithm applied to the Packer hypodermis dataset: (a) histogram of residuals, (b) Q-Q plot of residuals, (c) neighborhood graph, and (d) histogram of degrees



(a) Behavior of STAGE with respect to radius for Saelens data



(b) Behavior of STAGE with respect to radius for Packer seam cell data



(c) Behavior of STAGE with respect to radius for Packer hypodermis data

Figure 16. Behavior of STAGE with respect to neighborhood radius